

统计分析系列

内科科研统计设计与数据分析

胡良平 王素珍 郭 晋 主 编
吕军城 石福艳 任艳峰 副主编

電子工業出版社

Publishing House of Electronics Industry

北京 · BEIJING

内 容 简 介

本书作者查阅了多种内科学期刊,以每种期刊近5年中刊载的学术论文为研究对象,针对内科学期刊论文中统计学应用的现况和可能涉及的统计学知识,先以15章的篇幅从正面比较系统地介绍了开展内科学科研所必需的科研设计知识、统计表达描述知识、常用统计分析和多元统计分析知识;结合内科学论文中误用统计学的案例,介绍了“提高医学论文统计学质量的策略”和“内科学医学论文写作要领及应注意的问题”;介绍了遗传资料的统计分析与SAS实现;最后,还用4章篇幅提纲挈领地介绍了临床科研设计、数据探索性分析、统计分析方法合理选择以及临床科研统计学方面常见差错辨析与释疑要览。

本书适合从事生物医学、临床等工作的学生(大专生、本科生、硕士和博士研究生)、医护人员、科技工作者和学术杂志编审人员学习和使用。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

内科科研统计设计与数据分析/胡良平,王素珍,郭晋主编.—北京:电子工业出版社,2015.6
(统计分析系列)

ISBN 978-7-121-26270-8

I. ①内… II. ①胡… ②王… ③郭… III. ①统计学-应用-内科学-医学院校-教材 IV ①R5-05

中国版本图书馆CIP数据核字(2015)第125053号

策划编辑:秦淑灵

责任编辑:郝黎明 特约编辑:张燕虹

印 刷:

装 订:

出版发行:电子工业出版社

北京市海淀区万寿路173信箱 邮编 100036

开 本:787×1092 1/16 印张:24.5 字数:724千字

版 次:2015年6月第1版

印 次:2015年6月第1次印刷

册 数:3000册 定价:58.00元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zlt@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

《内科科研统计设计与数据分析》编委会

主 编 胡良平(军事医学科学院)

王素珍(潍坊医学院)

郭 晋(首都医科大学)

副主 编 吕军城(潍坊医学院)

石福艳(潍坊医学院)

任艳峰(潍坊医学院)

编 委 王小莉(军事医学科学院)

王 璐(军事医学科学院)

刘一松(军事医学科学院)

安洪庆(潍坊医学院)

李 双(军事医学科学院)

杜宗刚(潍坊医学院)

陈 娟(河北联合大学)

孟维静(潍坊医学院)

胡纯严(军事医学科学院)

贾玉凤(河北联合大学)

曹文君(长治医学院)

秘 书 赵晓蒙(潍坊医学院)

孔雨佳(潍坊医学院)

孙日扬(军事医学科学院)

刘惠刚(首都医科大学)

李子建(济南军区疾病预防控制中心)

吴庆文(河北联合大学)

李炳海(潍坊医学院)

李 崇(军事医学科学院)

胡 完(军事医学科学院)

郭春雪(军事医学科学院)

高 辉(解放军 95969 部队卫生队)

陶丽新(首都医科大学)

黄天资(潍坊医学院)

前 言

从事临床医学研究、特别是内科科研工作，不可避免地要运用统计学知识。因为导致人体患病的原因不计其数，而预防、诊断、治疗疾病和对治疗效果的评价又是十分复杂的。用统计学语言来概括：在医学研究领域里的任何角落，都充满了“多因素多指标且具有随机性”的问题与课题。而揭示大量随机变量变化规律必须依靠以下思想和技术：正确的统计思想、让理论密切联系实际三型理论、科学完善的科研设计方法、实时严格的质量控制和严密的系统统计理论与技术、基本常识和各学科领域专业知识，它们构成了医学统计学的完备体系。

如今，临床医生不仅要为广大民众预防疾病、为广大患者诊断和治疗疾病，要研究新的预防、诊断和治疗方法，还要不断总结临床经验和教训、正确评价疗效与安全性，从而不断提高临床研究和治疗水平；他们需要结合临床实际申请科研课题，并且需要发表高影响因子的 SCI 论文。所有这一切都表明：他们比以往任何时候都更加需要学习、掌握和正确运用统计学。本书正是为了满足这一现实需求应运而生的。

为了有效地帮助从事临床医学研究和一切需要统计设计与数据分析的人们正确领悟统计思想，把握统计学各部分的要领，将科研、教学、咨询等一切工作做得更加科学、严谨、经济、可靠，笔者与山东省潍坊医学院医学统计学教研室王素珍教授一道，组织多名统计学方面的年轻专家学者，针对多种内科学方面的学术期刊中出现的问题，概括地总结了各类期刊论文中统计学应用的现况。一方面，运用逆向思维，指出临床类学术期刊论文中存在的统计学错误；另一方面，又运用纵向思维，比较全面地介绍了高质量地完成一个临床研究课题应该具有的统计思想和统计技能，它们包括如何选题、如何构思、如何设计、如何进行质量控制、如何收集和整理资料、如何进行统计分析、如何撰写临床学术论文。

值得一提的是，笔者多年前的一位研究生郭晋，他在获得协和医科大学博士学位后刚入职首都医科大学不久的繁忙时期中，还挤出时间协助笔者审校了部分书稿，并为本书撰写了一章与遗传资料统计分析有关的内容；笔者另两位在读硕士研究生胡完和郭春雪，为本书中的每一章都做了认真且仔细的审读，纠正了许多原稿中的疏漏和不妥之处；另外，河北联合大学的吴庆文教授、军事医学科学院门诊部王小莉、王璐和李崇医生等，都为本书部分资料的收集与整理付出了大量辛勤的劳动；还有一些直接和间接为本书做出过贡献的人们。在此，一并致谢！

由于笔者水平有限，书中难免会出现这样或那样的不妥，甚至错误之处，恳请广大读者不吝赐教，以便再版时修正。

主编 胡良平
于北京军事医学科学院
生物医学统计学咨询中心
2015 年 2 月

目 录

第 1 章 内科科研的思路与要领	(1)
1.1 什么是内科科研	(1)
1.2 内科科研的流程	(1)
1.2.1 提出内科研究问题阶段	(1)
1.2.2 内科科研设计阶段	(3)
1.2.3 内科科研实施阶段	(4)
1.2.4 内科科研数据处理阶段	(4)
1.2.5 总结撰写论文	(5)
1.2.6 内科研究的分类	(5)
1.3 内科研究方法	(7)
1.3.1 选题内容	(8)
1.3.2 研究方法及统计学应用	(12)
1.4 如何学习统计学	(14)
1.4.1 统计学的整体构架	(14)
1.4.2 正确的统计思想	(14)
1.4.3 三型理论	(15)
1.4.4 学好统计学的要领	(15)
参考文献	(15)
第 2 章 常用内科统计设计与数据分析	(16)
2.1 单因素 K 水平 ($K \geq 2$) 设计与分析	(16)
2.1.1 应用目的	(16)
2.1.2 设计方法	(16)
2.1.3 具体设计过程	(16)
2.1.4 统计分析方法	(17)
2.1.5 SAS 实现程序及结果解释	(17)
2.1.6 实例的辨析与释疑	(21)
2.2 配对设计与分析	(25)
2.2.1 应用目的	(25)
2.2.2 设计方法	(25)
2.2.3 具体设计过程	(25)
2.2.4 统计分析方法	(26)
2.2.5 SAS 实现程序及结果解释	(26)
2.3 随机区组设计与分析	(28)
2.3.1 应用目的	(28)
2.3.2 设计方法	(28)

2.3.3	具体设计过程	(29)
2.3.4	统计分析方法	(29)
2.3.5	SAS 实现程序及结果解释	(29)
2.3.6	实例辨析	(32)
2.4	析因设计及其一元定量资料的统计分析	(33)
2.4.1	应用目的	(33)
2.4.2	设计方法	(33)
2.4.3	具体设计过程	(34)
2.4.4	统计分析方法	(34)
2.4.5	SAS 实现程序及结果解释	(34)
2.5	正交设计	(39)
2.5.1	应用目的	(39)
2.5.2	设计方法	(40)
2.5.3	具体设计过程	(40)
2.5.4	统计分析方法	(41)
2.5.5	SAS 实现程序及结果解释	(41)
2.5.6	实例辨析与释疑	(45)
2.6	拉丁方设计	(46)
2.6.1	应用目的	(46)
2.6.2	设计方法	(46)
2.6.3	具体设计过程	(47)
2.6.4	统计分析方法	(48)
2.6.5	实例辨析与释疑	(50)
2.7	交叉设计与分析	(50)
2.7.1	设计方法	(50)
2.7.2	具体设计过程	(50)
2.7.3	SAS 实现程序及结果解释	(51)
2.8	重复测量设计	(52)
2.8.1	设计方法	(52)
2.8.2	具体设计过程	(52)
2.8.3	统计分析方法	(53)
2.8.4	SAS 实现程序及结果解释	(53)
2.8.5	实例辨析与释疑	(56)
第3章	内科试验设计与数据分析	(58)
3.1	内科试验设计的三要素	(58)
3.1.1	内科试验设计概述	(58)
3.1.2	试验设计的三要素	(58)
3.2	试验设计的基本原则	(62)
3.2.1	随机化原则	(62)
3.2.2	对照原则	(62)
3.2.3	重复原则	(65)
3.3	内科试验设计方法	(66)

3.3.1	剂量设计的方法	(68)
3.3.2	试验设计方法	(69)
3.4	动物试验常见问题及其对策	(78)
3.4.1	我国动物试验领域存在的问题	(78)
3.4.2	我国动物试验替代方法(3R)研究存在问题	(79)
3.4.3	针对我国动物试验问题的对策	(79)
附例1	内科学动物试验设计主要包括以下基本步骤	(80)
附例2	诱发性肿瘤动物试验设计	(81)
附例3	著名医学科学家的重大发现与动物试验的关系	(82)
第4章	内科临床试验设计与数据分析	(87)
4.1	内科临床试验的基础知识	(87)
4.1.1	临床试验的定义	(87)
4.1.2	临床治疗试验的特点	(87)
4.1.3	临床试验的分期	(88)
4.1.4	临床试验的统计分析集	(88)
4.2	内科临床试验设计的要素及原则	(89)
4.2.1	内科学试验设计的三要素	(89)
4.2.2	内科学试验设计的基本原则	(92)
4.2.3	科研设计中样本含量的确定	(104)
4.2.4	内科学试验设计中实例展示	(111)
4.3	临床试验的质量控制	(113)
4.3.1	试验误差的种类	(113)
4.3.2	误差的控制	(114)
4.4	随机双盲临床试验的实施过程	(114)
4.4.1	随机双盲临床试验实施过程	(114)
4.4.2	双盲临床试验中须注意的问题和技巧	(116)
4.5	临床试验实例	(117)
4.5.1	研究设计	(118)
4.5.2	样本含量估计	(118)
4.5.3	病例选择	(118)
4.5.4	给药方案	(120)
4.5.5	疗效指标	(120)
4.5.6	数据录入与核查	(121)
4.5.7	统计分析	(121)
4.6	诊断试验的评价	(122)
4.6.1	诊断试验的设计步骤	(123)
4.6.2	诊断试验的评价指标	(123)
4.6.3	联合试验	(125)
4.6.4	受试者工作特征曲线	(126)
4.6.5	诊断试验临界点的确立方法	(127)
4.6.6	诊断试验评价中常见的偏倚	(128)

第5章	内科科研调查设计与数据分析	(129)
5.1	调查设计的基本内容	(129)
5.1.1	调查的概念及其特点	(129)
5.1.2	调查设计的基本原则与内容	(129)
5.1.3	标准化率	(132)
5.1.4	调查表和问卷的设计与评价	(133)
5.1.5	内科调查设计样例展示及对差错的辨析与释疑	(137)
5.2	横断面调查研究的设计与分析	(138)
5.2.1	调查方法	(138)
5.2.2	常用的抽样方法	(139)
5.2.3	样本含量的估计	(144)
5.2.4	资料的整理和分析	(145)
5.2.5	内科横断面调查研究样例展示及对差错的辨析与释疑	(145)
5.3	追踪性调查研究的设计与分析	(146)
5.3.1	基本概念	(146)
5.3.2	追踪性调查研究的类型	(147)
5.3.3	研究对象的选择	(148)
5.3.4	样本含量的估计	(149)
5.3.5	资料的整理和分析	(149)
5.3.6	追踪性调查研究中的偏倚及控制	(152)
5.3.7	追踪研究的优缺点	(153)
5.3.8	内科追踪性调查研究样例展示及对差错的辨析与释疑	(153)
5.4	回顾性调查研究的设计与分析	(154)
5.4.1	基本概念	(154)
5.4.2	回顾性调查研究的类型	(155)
5.4.3	病例与对照的选择	(155)
5.4.4	样本含量的估计	(156)
5.4.5	资料的分析	(157)
5.4.6	回顾性研究中的偏倚及控制	(159)
5.4.7	回顾性研究的优缺点	(160)
5.4.8	内科学回顾性调查研究样例展示及对差错的辨析与释疑	(160)
第6章	常用内科科研数据统计分析方法	(163)
6.1	如何合理应用统计描述指标	(163)
6.1.1	统计资料的类型	(163)
6.1.2	如何合理应用统计描述指标描述定量资料	(163)
6.1.3	如何合理应用统计描述指标描述定性资料	(168)
6.2	如何合理应用统计图、统计表	(171)
6.2.1	统计表	(171)
6.2.2	统计图	(174)
6.3	如何合理选用统计分析方法	(180)
6.3.1	定量资料统计分析方法的合理选用	(180)

6.3.2	定性资料统计分析方法的合理选用	(186)
6.4	如何合理应用相关和回归分析	(205)
6.4.1	散点图呈直线趋势, 方可做直线相关和回归分析	(205)
6.4.2	对关联性强度定位错误	(206)
6.4.3	滥用直线相关分析得出没有实际意义的结论	(206)
6.4.4	误用 t 检验方法说明相关回归问题	(206)
6.4.5	误用直线相关分析代替 Spearman 等级相关分析	(207)
6.4.6	误用卡方检验来说明相关回归才能回答的问题	(208)
6.4.7	多重线性回归分析的合理应用	(210)
6.5	如何合理应用诊断试验评价指标	(210)
6.5.1	常用诊断试验评价指标	(211)
6.5.2	诊断试验的正确应用	(212)
6.5.3	诊断试验常见错误及错误辨析	(212)
6.6	如何合理应用 ROC 曲线分析	(213)
6.6.1	ROC 曲线简介	(213)
6.6.2	ROC 曲线下面积计算	(214)
6.6.3	ROC 曲线分析的 SAS 实现	(215)
第 7 章	协方差分析	(218)
7.1	协变量与协变量控制的基本思想	(218)
7.1.1	协变量的概念	(218)
7.1.2	协方差分析的适用条件	(218)
7.2	协方差分析的方法与步骤	(219)
第 8 章	多重线性回归分析	(227)
8.1	多重线性回归方程的建立及假设检验	(227)
8.1.1	多重线性回归模型	(227)
8.1.2	多重线性回归方程的建立及假设检验	(228)
8.1.3	多重回归方程的评价	(229)
8.2	多重回归分析中自变量的筛选	(230)
8.2.1	全局择优法	(230)
8.2.2	局部选择法	(231)
8.3	多重回归分析应用的注意事项	(232)
8.3.1	模型的选择	(232)
8.3.2	样本含量	(233)
8.3.3	指标的数量化	(233)
8.3.4	多重线性回归分析中的多重共线性问题和交互作用	(233)
8.3.5	残差分析	(234)
8.4	案例分析	(235)
第 9 章	Poisson 回归模型	(241)
9.1	Poisson 回归模型及参数估计	(241)
9.1.1	模型拟合	(241)

9.1.2	参数检验	(241)
9.2	实例分析	(241)
第 10 章	logistic 回归	(244)
10.1	logistic 回归	(244)
10.1.1	二分类反应变量的 logistic 回归分析	(244)
10.1.2	logistic 回归模型的参数估计和假设检验	(245)
10.1.3	logistic 回归模型的流行病学意义	(248)
10.2	条件 logistic 回归	(249)
10.2.1	基本原理	(249)
10.2.2	参数估计与假设检验	(250)
10.3	logistic 回归分析的应用及其注意事项	(250)
10.3.1	logistic 回归分析的应用	(250)
10.3.2	logistic 回归分析的注意事项	(251)
10.4	案例分析	(253)
第 11 章	生存分析	(259)
11.1	生存分析中的基本概念	(259)
11.1.1	基本概念	(259)
11.1.2	资料收集	(261)
11.1.3	生存分析的基本方法	(261)
11.2	生存率的估计与生存曲线	(262)
11.2.1	小样本资料生存率及其标准误的计算	(262)
11.2.2	大样本资料的生存分析	(263)
11.3	生存曲线的假设检验	(264)
11.3.1	log-rank 检验	(264)
11.3.2	趋势检验	(264)
11.3.3	Breslow 检验	(264)
11.4	Cox 比例风险模型	(265)
11.4.1	Cox 回归模型的构造	(265)
11.4.2	Cox 回归模型的拟合步骤	(266)
11.4.3	偏似然函数及参数估计的计算方法	(266)
11.4.4	回归系数的统计学意义检验	(267)
11.5	三种常见回归模型的选择	(267)
11.6	案例分析	(268)
11.7	生存分析案例辨析	(275)
第 12 章	判别分析	(277)
12.1	判别分析的基本思想	(277)
12.2	判别分析的步骤	(277)
12.3	实例分析	(278)
第 13 章	聚类分析	(281)
13.1	聚类分析的基本思想	(281)

13.2	样品系统聚类分析步骤	(282)
13.3	指标分解聚类分析步骤	(282)
13.4	实例分析	(282)
13.4.1	问题与数据	(282)
13.4.2	对给定资料进行无序样品聚类分析	(283)
13.4.3	对给定资料进行变量聚类分析	(285)
第 14 章	主成分分析	(289)
14.1	主成分分析的基本原理	(289)
14.2	主成分的贡献率	(290)
14.3	主成分分析的步骤	(290)
14.4	实例辨析	(291)
	参考文献	(293)
第 15 章	探索性因子分析	(294)
15.1	因子分析模型	(294)
15.2	因子分析的步骤	(294)
15.3	实例分析	(295)
第 16 章	遗传数据统计分析	(300)
16.1	基因、基因型频率测定与 Hardy-Weinberg 平衡定律的验证	(300)
16.2	连锁不平衡与单体型分析	(303)
16.2.1	概述	(303)
16.2.2	最大似然法估计单体型概率	(303)
16.2.3	E-M 算法估计单体型概率及对连锁不平衡进行假设检验	(303)
16.3	多位点基因型与疾病关联分析	(306)
16.4	标签 SNP 的确认与 SAS 程序	(308)
16.5	一般人群病例对照遗传资料的关联分析	(310)
16.6	家系数据的关联分析	(312)
16.6.1	父母亲作对照的关联分析原理介绍	(312)
16.6.2	同胞作对照的关联分析原理介绍	(313)
16.7	本章小结	(316)
第 17 章	提高医学论文统计学质量的策略	(317)
17.1	设计阶段存在的问题	(317)
17.2	数据分析阶段存在的问题	(320)
17.2.1	常被误用的统计检验方法	(321)
17.2.2	使用卡方检验人们常犯的典型错误	(321)
17.2.3	数据分析阶段有问题的样例展示及对差错的辨析与释疑	(321)
17.3	资料整理阶段常出现的问题	(323)
17.4	数据呈现阶段常出现的问题	(324)
17.5	数据解释中出现的错误	(324)

第 18 章	内科论文写作要领及注意事项	(326)
18.1	内科论文常见的写作问题	(326)
18.1.1	题目常见写作问题	(326)
18.1.2	摘要常见写作问题	(326)
18.1.3	表格常见写作问题	(326)
18.2	内科论文写作要求及注意事项	(328)
18.2.1	写作要求	(328)
18.2.2	注意事项	(330)
	参考文献	(333)
第 19 章	临床科研统计设计要览	(334)
19.1	临床科研的种类	(334)
19.1.1	调查某些疾病的发病率或人群分布规律	(334)
19.1.2	为某些医学指标制定参考值范围或为某些疾病的诊断与治疗制定标准	(334)
19.1.3	为评价某治疗方法或药物或医疗器械的有效性与安全性	(335)
19.1.4	同类临床研究的 meta 分析	(335)
19.1.5	为探查某些疾病的发病原因的研究	(335)
19.1.6	基因治疗研究	(335)
19.2	临床科研的核心内容	(336)
19.2.1	临床科研需要关注有关的政策和法规	(336)
19.2.2	临床科研中不可回避的伦理道德问题	(336)
19.2.3	临床科研中应尽可能降低心理因素对结果的影响	(337)
19.2.4	临床科研中应尽可能提高受试者的依从性	(337)
19.2.5	临床研究中应全面把握试验设计的三要素	(338)
19.2.6	临床研究中应认真落实四原则	(338)
19.2.7	临床研究中应合理选用设计与比较类型	(338)
19.2.8	实时并严格控制临床试验研究质量	(339)
19.2.9	正确无误地在临床病例报告表上进行记录	(340)
第 20 章	临床科研资料统计表达描述要览	(341)
20.1	如何正确地运用统计表	(341)
20.1.1	常见统计表的种类	(341)
20.1.2	常见统计表举例	(341)
20.2	如何以数据库的形式呈现多因素多指标的复杂临床科研资料	(345)
20.2.1	概述	(345)
20.2.2	以数据库的形式呈现多因素多指标的复杂临床科研资料	(346)
20.3	如何合理地选用统计图	(348)
20.3.1	概述	(348)
20.3.2	合理选择统计图表达定性资料	(348)
第 21 章	临床科研资料统计分析要览	(352)
21.1	统计分析方法的分类	(352)
21.1.1	概述	(352)

21.1.2	探索性统计分析	(353)
21.1.3	差异性统计分析	(353)
21.1.4	相关与回归分析	(353)
21.1.5	广义综合评价	(356)
21.2	合理选用统计分析方法的关键技术	(358)
21.2.1	合理选用统计分析方法的四要素	(358)
21.2.2	合理选用统计分析方法的实例演示	(359)
第 22 章	临床科研统计设计与数据分析常见差错辨析与释疑要览	(361)
22.1	康复医学研究论文统计学差错辨析与释疑	(361)
22.1.1	基于多重线性回归分析的临床试验研究	(361)
22.1.2	基于具有一个重复测量因素的两因素设计的临床试验研究	(363)
22.2	非康复医学研究论文统计学差错辨析与释疑	(364)
22.2.1	基于单因素 4 水平设计的临床试验研究	(364)
22.2.2	基于生存分析的临床试验研究	(365)
附录 A	胡良平统计学专著及配套软件简介	(368)

第1章 内科科研的思路与要领

科研设计是医护工作者开展科学研究的重要内容。医护工作者从事内科科研是一项创新性的工作，是推动内科学学科发展，促进内科学理论、知识、技能更新的有效措施。随着内科学学科的发展，内科科研日益受到广大内科学工作者的重视。合理的医学研究设计与分析在内科科研工作中起着重要作用。任何一项研究，首先要确定研究目的，之后并不是立即付诸实施，而是首先考虑如何做才可能达到预期的目的，实施过程中如何确保研究质量，可能得到怎样的结果，将如何处理所获得的资料。将解答前述内容所形成的书面材料称为课题的统计设计方案。科学完善的课题统计设计是内科研究成果符合科学性要求的重要保证，是所得结果和结论经得起时间和实践检验的重要前提。

1.1 什么是内科科研

在医学上，习惯上分为“内、外、妇、儿”等科。所谓内科，就是医疗机构中主要用药物而不用手术来治疗内脏疾病的一科。从宏观上来讲，内科研究是围绕内科领域中预防、诊断、治疗或预后评价等方面不断地提出问题和解决问题的探索过程，每一个科研课题首先都必须是一个“问题”。所谓“问题”，就是人们想要解决而没有解决的难点，是知和不知的对立统一。它体现了人们对事物的已有认识和没有认识之间的矛盾。如果对内科本身一无所知，就形不成内科研究问题。

从微观上来讲，内科科研是指研究者提出内科领域中的理论与实践问题，并采用科学的方法反复实践、观察、探索，系统地研究和解决该问题，并直接或间接地促进内科学理论的发展和指导内科学实践过程。内科科研的意义在于丰富和完善内科学的理论体系，推动内科学理论与实践的结合，提高内科研究工作的水平和质量，也有利于内科学工作者素质和技能的培养与提高。

1.2 内科科研的流程

内科科研过程是一个提出假说并验证假说的过程，这一过程一般包括以下几个基本阶段，即提出内科研究问题阶段、设计阶段、实施阶段和数据处理阶段等。

1.2.1 提出内科研究问题阶段

一般来说，该阶段要经过以下4个步骤。

1. 提出内科问题，形成最初的意念

这是科学研究工作的必要准备，又是科学研究工作的一个重要组成部分。选择内科研究的主攻方向就是内科研究者(集体或个人)首先要确定的在一个较长时间内进行科学攻关的大方向。内科研究主攻方向下面又包含着一系列的既相互联系、又相互区别的具体科研课题，它是一个课题群。确定研究课题也就是研究者在科学研究过程中的第一阶段，从课题群中筛选并确定一个(或几个)具体研究课题。这是最重要的一步，它决定整个研究设计和研究工作的全过程。通常，内科研究的选题多来源于：(1)从他人的研究结果中获得启发；(2)个人实践经验和日常工作；(3)内科学发展中需要探讨的内容；(4)研究者的兴趣爱好；(5)从内科研究项目指南中选题。

一个好的选题不应凭空杜撰，必须遵守以下几条共性原则。

(1) 选题要有明确的目的，应根据内科学当前发展的实际，解决内科学目前存在的实际问题，选取具有重大意义或迫切需要解决的课题来进行研究。目标应集中，不含糊、不笼统。

(2) 创新性原则，即价值原则。就是指内科选题要有新颖性、先进性，要选择前人没有解决或没有完全解决的课题进行研究，不能只重复前人做过的工作。坚持创新，就是要善于捕捉有价值的线索，勇于探索，不断深化。创新可分为两种类型：根本性创新和增量性创新。主要是看选题的内容是否开拓新领域，提出新思想、新理论；是否采用了新设计、新方法等。要有所发明、有所发现，其学术水平应有所提高，要能够推动内科学学科向前发展。

(3) 科学性原则。从选题来看，必须有充分理论依据，即要有理论基础，理论基础来自前人的经验总结和个人研究工作的实践；要符合客观规律，从实际出发，实事求是，尊重客观规律；设计要科学，符合逻辑，应该采用专业和统计两个方面的知识对课题进行设计，保证受试对象对总体有极好的代表性（有依据地估计出足够的样本量，并通过严格遵守随机、均衡原则来获取和分配受试对象）和效应结果的可重复性。研究结果要经得起推敲、重复，经得起时间、实践和历史的检验。

(4) 实用性原则。即能够应用于当前社会的医疗发展实践中，最好能转化为生产力，成为促进经济和社会发展的动力，并能够在实践中不断经受检验并使之进一步得到发展，这是它的社会意义；从内科学本身发展的需要来看，实用性是指课题的研究成果应该能够与临床实践相结合，研究疾病的预防、诊断、治疗和康复，最大限度地为患者减轻痛苦。

(5) 可行性原则，指课题的主要内容和技术指标实现的可能性。除设计方案和技术路线科学、可行外，以下条件也是影响课题可行性的重要因素：课题负责人的学术水平；课题组的专业、知识和年龄结构；仪器、设备、信息、试剂和实验动物；做好预实验的程度等。选题要从实际出发，脚踏实地，结合本人、本单位客观条件，量力而行。

(6) 预期成果：一项好的研究课题，如果预期结果不明确，或根本达不到，或者周期很长，都不是好课题。

2. 查阅文献

虽然在课题选择或内科临床观察之前，已对有关资料和学术动态进行了搜集和分析，但是在题目确定以后，仍要查阅大量有关文献，以作为对已掌握的文献的补充。有人做过统计，国内外多数科学工作者查阅文献的时间约占整个科研工作的三分之一。如果没有这些最新的参考文献，要想使内科研究达到新颖和独创性是不可能的。由此可见，查阅搜集文献在整个科研过程中的重要性及必要性。现在文献查阅已非常方便，可以通过因特网(Internet)直接检索。一般来说，对某一内科问题，国内外都没有研究的可能性不大，或多或少都能检索到相关的研究文献。通过查阅文献，可以对所立项目的新颖性、所选择方法的先进性和科学性进行自我评价；通过文献查阅，还可以为课题的研究开拓思路，提供理论依据。因此在查阅文献时，应根据内科课题的需要，从学报、论文集、百科全书、教科书、专著等资料中查阅文献，把与课题有密切关系的资料做好记录，注明文献的出处。

3. 形成内科学假说

假说是在初始意念和文献查阅的基础上，以已知的科学事实为依据，进行科学总结、推测而形成的，是初始意念的深化、系统化。著名的微生物学家巴斯德曾说：“有科学的假说，才能有有价值的观察。”假说就是科研思路(IDEA)。1987年诺贝尔物理奖获得者——卡尔·缪勒(瑞士)曾说：研究者必须要有好奇心、功名心，特殊的才智、特殊的记忆力、工作的纪律性，恰如其分地估价自己。科研思路是创造性劳动、是经验的再发展、是知识的拓宽。IDEA需要灵感，更需要工作基础，是能力的呈现。因此，IDEA有知识产权，也是你为之奋斗的起点。Doll和Hill(1950—1964)研究吸烟

与肺癌的关系，起源于他们的仔细观察和用心思考，他们根据英国扫烟囱工人多发阴囊性皮癌，接触苯胺染料的工人易患膀胱癌，辐射可以引发骨肉瘤等一系列事实，经过归纳、推理，认为某些化学物质可以致肿瘤。因此，设想烟草的不完全燃烧也可以产生致癌物质，促使肺癌发生，提出了吸烟可以致癌的假说。

4. 陈述问题

陈述问题主要是说明选题的依据、背景及进展，选题的特点及创新之处，假说的形成过程及其内容，拟采取的实验手段、工作进程、预期结果，以及课题的水平和学术、应用价值等。

1.2.2 内科科研设计阶段

设计阶段是连接陈述与实施阶段的桥梁，该阶段要围绕如何验证假说或提出的问题制定一个实施方案，即课题统计设计方案。比如在 Doll 和 Hill (1950—1964) 对吸烟与肺癌的关系的研究中，在提出吸烟可以致癌假说的前提下，综合运用演绎推理，安排了各项调查研究，对假说进行了验证，获得了成功。科研设计包括专业设计和统计设计两个方面。

专业设计是根据假说运用内科专业知识拟定研究内容、对象、方法、实验方案和技术路线的过程，是保证内科科研课题先进性和有用性的重要手段。比如苦瓜冲剂治疗糖尿病的研究：第一，要从专业角度进行设计，包括设计研究内容，在此即通过临床试验验证苦瓜冲剂的疗效，需要选择受试对象，选定糖尿病患者作为病例；第二，确定试验因素，如苦瓜冲剂的不同剂量；第三，选择试验方法，即采用临床试验方法，给予糖尿病患者不同剂量的苦瓜冲剂治疗；第四，观测试验效应，即通过临床试验的观察和检测指标的数值大小来反映，是判断试验结果的尺度，具体就是糖尿病患者的血糖值和糖化血红蛋白值等的改变情况。

统计设计则是围绕专业设计，确定统计设计类型、样本大小、分组方法、统计分析指标及统计分析方法等，实际是运用数理统计理论和方法来进行设计。减少抽样误差和排除系统误差，保证样本的代表性和样本间的可比性，确保实验观察内容的合理安排，以便对实验结果进行高效率的统计分析，以最少的实验观察次数(例数)得出相对最优的结果和可靠的结论。因此，统计设计是科研成果可靠性和经济性的保证。统计设计要遵循对照、随机、重复和均衡的原则。同样，以上述不同剂量苦瓜冲剂治疗糖尿病的研究为例，进行统计设计与专业设计是有区别的。在统计设计中，首先要确定一个或多个具体的设计类型来安排因素及其水平。比如是完全随机设计、随机区组设计还是析因设计或重复测量设计等。好的开始是成功的一半，设计类型选择正确，是后续工作顺利实施的保障。设计类型一旦确定，下一步是根据设计类型和已知的相关指标等信息计算合适的样本量。样本量的确定也是非常重要的，样本量过小会得出假阴性结论，过大又会导致浪费，应该根据专业知识和相应的指标计算出合适的样本量。统计设计要遵循对照、随机、重复和均衡原则，因此要为研究的药物设定一个对照，即在让一组患者服用苦瓜冲剂的同时，设定与该组对等的另一组服用阳性药物或者安慰剂作为对照；如果采用的设计类型是完全随机设计的，则需要将糖尿病患者随机分为两组，一组服用苦瓜冲剂，另一组服用阳性药物或安慰剂；当然，也可以采用平行对照，即 $K(K \geq 2)$ 组糖尿病患者分别服用不同剂量的苦瓜冲剂。这其实也是确定分组的方法，分组的时候一定要遵循随机化原则，以保证组间的可比性；遵循重复性原则就是保证实验要有一定的样本含量，任何药物的疗效都不能通过仅有的几个病例来说明，样本要达到一定的数量才可以(因为特定剂量的药物的疗效在少数患者身上的表现是随机的，而在大量患者身上的表现就趋于稳定了)；确定分析指标，在这里与专业的设计相同，比如在上述实例中，确定血糖或者糖化血红蛋白等作为主要疗效评价指标，确定尿糖作为次要疗效的评价指标等。最后一步是进行统计分析，统计分析取决于设计类型，比如如果采用两组完全随机设计，则可以采用单因素两水平设计一元定量资料 t 检验或者采

用方差分析(定量资料必须满足三个前提条件,即独立性、正态性和方差齐性)或相应设计一元定量资料秩和检验;如果是配对设计,则可以采用配对设计一元定量资料 t 检验(前提条件为独立性和正态性)或配对设计一元定量资料符号秩和检验对血糖或者糖化血红蛋白等定量指标的测定结果进行统计分析。

总之,专业设计和统计设计都是科研设计的两个重要组成部分,二者相辅相成,缺一不可。

1.2.3 内科科研实施阶段

该阶段即按照设计方案进行研究,搜集论证假说的证据,积累数据资料的过程,也是内科研究中重要的一步。内科研究既可以是动物实验,也可以是临床试验。无论是哪种,都应根据设计方案,实施实验研究并严格控制实施过程中的质量,通过观察实验结果搜集资料。该过程中,应该严格按照设计要求、实验步骤和操作规程进行实验,切实保证实验的数量和质量要求。要认真观察实验中的各种反应,真实记录实验中的阴性、阳性结果,确保实验的准确性、可靠性和可重复性。

记录数据要准确、客观,严防主观臆造,也不可任意去除实验中的任何阴性反应,记录的过程中善于分析对比。不能“各取所需”,或者废弃与自己主观愿望不一致的数据、资料。

先进的方法通常是获得先进水平的手段,特别是临床试验研究,应注意采用先进的技术和设备获取资料,注重与先进水平相联系,依靠高科技和熟练的操作技术,获得稳定可靠的试验数据,减少试验误差,这也是获得具有科研价值的研究成果的基础。因此,科学地搜集、获取资料是积累经验、发现规律、提高认识、总结并发展理论的前提,是科研成功的基础。

1.2.4 内科科研数据处理阶段

科研数据的处理即统计分析的过程,是利用观察和实验所搜集的数据进行整理和加工,分组运算,在此基础上进一步对资料进行描述、推断、比较、分析、估计、评价、预测等处理,使所获得的资料更加系统、内在的规律更加显而易见,从而得出科学结论的过程。正确的统计分析方法可以帮助内科科研人员采用正确和理性的思维方式,排除偶然、发现必然;透过现象发现本质,通过事物的数量关系而揭示事物间的内在联系,发现其内在的规律性。统计处理可以令研究成果更加科学合理,是内科学研究中的重要步骤。

在统计学中,人们将资料分为定量资料和定性资料。研究者从观察对象身上得到某个指标的具体取值,如身高、体重、血压等,这样的资料被称为定量资料;有的指标没有具体的取值,只能用一种性质来描述。如血型、性别、民族等,这样的资料被称为定性资料。不管是定量资料还是定性资料,它们的统计分析都包括统计描述和统计推断。

统计描述就是采用合适的统计指标描述资料的性质和特征。应该根据资料的不同性质、不同特点合理选择。比如定量资料可以通过集中趋势和离散程度的指标来描述,集中趋势指标包括算术均数、几何均数、中位数等;离散程度指标包括极差、四分位数间距、方差、标准差、变异系数等;应根据数据的类型与分布选择正确的统计指标。例如同样是对集中趋势进行描述,计算平均年龄时常选择算术均数,而乙肝的平均潜伏期则常选中位数。定性资料可以用相对数、绝对数来描述,比如率、相对比等。此外,还可以用统计表和统计图等来表达所整理的资料。

统计推断就是根据现有的样本信息推断总体的特征。在推断时需针对不同设计类型下的资料与分布特征正确应用统计分析方法,即使是相同类型的资料,不同设计下的分析方法也各不相同。例如方差分析可用于不同设计类型的定量资料,包括完全随机设计、随机区组设计、析因设计、重复测量设计等。应根据设计类型和资料性质以及研究目的,合理选择统计分析方法。

1.2.5 总结撰写论文

在研究的最后阶段，应以文字形式对研究工作进行总结。论文的撰写和发表是进行研究工作小结和学术交流的平台。论文是对整个医学科研工作的概括，但又不同于一般的工作总结报告，是科学研究实践的理论升华。论文撰写应遵循科学性、先进性、实用性及严肃性的原则，应该有明确的论点、充分的科学论据和系统的论证过程，应该在充足的证据和科学的推理之上，客观、准确、实事求是地表达研究的过程和结果。国际医学期刊编辑委员会(ICMJE)发布的《向生物医学期刊投稿的统一要求》(Uniform Requirements for Manuscripts Submitted to Biomedical Journals)，包含了对医学科研论文写作不同阶段的要求。此“统一要求”提供的是适用于所有期刊稿件的普遍性指南，具体到某个期刊的特殊要求，一般都会在投稿须知中提供相应的信息。论文由论文题目、作者、摘要、关键词、前言、研究对象、材料与方法、结果、讨论、致谢、参考文献等部分组成。各个部分或多或少都会涉及统计设计和统计结果的表达及解释。

内科科研的基本流程可用图 1-1 来表示。其中，内科统计设计、资料与整理和统计分析这三个阶段都需要运用统计学知识，可见统计学对内科科研的重要性。

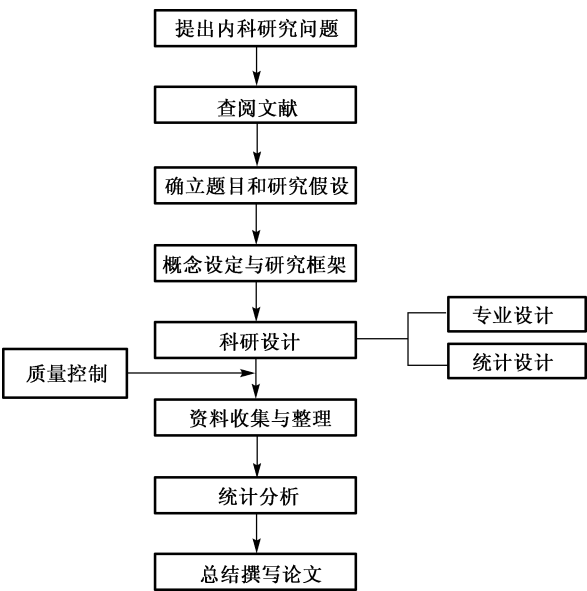


图 1-1 内科科研的基本流程

1.2.6 内科研究的分类

1. 依据研究目的分类

依据研究目的可划分为基础研究(fundamental research)、应用研究(applied research)和开发研究(development research)。

基础研究指为获得关于内科现象和可观察事实的基本原理及新知识而进行的实验性和理论性工作，它不以任何专门或特定的应用或使用为目的。主要目的是增加科学技术知识和发现探索领域。一般不考虑近期实用目标，看不到或无明显社会效益或经济效益的研究。

基础研究的主要特点：(1)以认识现象、发现和开拓新的知识领域为目的，即通过实验分析或理论性研究对事物的物性、结构和各种关系进行分析，加深对客观事物的认识，解释现象的本质，

揭示物质运动的规律,或者提出和验证各种设想、理论或定律;(2)没有任何特定的应用或使用目的,在进行研究时对其成果看不出、说不清有什么用处,或虽肯定会有用途但并不确知达到应用目的的技术途径和方法;(3)一般由科学家承担,他们在确定研究专题以及安排工作上有很程度的自由;(4)研究结果通常具有一般的或普遍的正确性,成果常表现为一般的原则、理论或规律并以论文的形式在科学期刊上发表或学术会议上交流。

内科中的基础研究范围主要包括:(1)保持人体健康的规律,健康指标的分期的基础研究;(2)人体功能与结构的研究,比如研究正常人体器官内的神经走行、血管在某器官内的分布;(3)疾病发生、发展、转归全过程的规律及分子基础,比如动脉粥样硬化机理研究、矽肺发生机理研究、冠心病患者血液中某些物质含量的变化;生长因子促进增殖型糖尿病毛细血管新生作用研究;复方丹参片对肾小球动脉硬化化的影响;(4)人体衰老过程的规律及分子基础,DNA 损伤与修复过程中基因结构的变化、T 细胞 E 受体受刺激后细胞内生化的变化等;(5)人体的生物力学、流体力学、电子学,比如心律失常的电生理特性及其机制;(6)化学药物的构效关系,植物药的亲缘与有效成分关系等,比如鱿鱼抗肿瘤成分的分离纯化与理化性质分析;(7)经络本质的研究等,这些都是基础研究。

应用研究(applied research):指为获得新知识而进行的创造性研究,它主要是针对某一特定的实际应用目的或目标,既是发展医学科研基础研究的成果,又是为达到某些特定的和预先确定的实际目标提供新的方法或途径。其成果对科学技术领域的影响是有限的,针对性更强。在内科学中,主要表现在以下几个方面:(1)有关疾病的病因、流行规律、治疗及防治方法的机制研究;(2)为实验研究需建立的新的动物模型、细胞株以及方法学的研究;(3)有关流行病学调查、考核防治效果、药物调查的方法学研究;(4)寻找新药物、新生物制品、新医用材料的方法、有效药物的药理作用机制、药代动力学、医用材料的机体相容性的机制研究等。

应用研究特点:(1)具有特定的实际目的或应用目标,具体表现是,为了确定基础研究成果可能的用途或为达到预定的目标探索应采取的新方法(原理性)或新途径;(2)在围绕特定目的或目标进行的过程中获取新的知识,为解决实际问题提供科学依据;(3)研究结果一般只影响科学技术的有限范围,并具有专门的性质,针对具体的领域、问题或情况,其成果形式以科学论文、专著、原理性模型或发明专利为主。简单地说,所谓应用研究,就是将理论发展成为实际运用的形式。

内科学中的应用研究也非常常见。比如脐血干细胞移植治疗粒细胞白血病;不同抗精子膜蛋白抗体对抗生育作用影响的比较研究;网织红细胞与骨髓瘤细胞杂交系的染色体组型、分带与恶性肿瘤的关系;建立临床检测 T 细胞功能的技术方法;下咽及颈段食管癌切除后硅橡胶管重建的研究;新光辐射治疗剂的有效成分的生物学活性及其作用原理,SARS 疫苗的研制与应用研究等都是应用研究。

开发研究(development research):开发研究是利用应用研究的成果和现在的知识与技术,创造新技术、新方法和新产品,是一种以生产新产品或完成工程技术任务为内容而进行的研究活动。在内科研究中,其主要内容包括:(1)有关疾病的新型诊断、治疗、预防方法及措施的研究;(2)有关新药物、新生物制品、新仪器器械、新试剂、新医用材料实验室样品的研制;(3)有关药物的资源调查、植物药的引种实验等。

以下几个例子是内科学中常见的应用研究:心律失常的药物治​​疗及应用起搏器手术的指征;异常血红蛋白病的产前诊断;肢体动脉硬化闭塞病的无创性测定技术的研究;降血脂新药德欧欣捷宁胶囊的临床研究;第二代国产肾衰专用必需氨基酸的试制和临床观察;脉冲激光器的研制和改进。

由以上的介绍可以看出,基础研究、应用研究的主要标志是增加科学、技术知识,以直接产生社会效益为主,而开发研究的标志是推广应用新技术、新产品,以直接产生经济效益为主。

2. 依据研究内容和方法分类

按照研究的内容和方法可分为调查研究和实验研究。

(1) 调查研究: 研究者为了解人群的健康状况(疾病的分布、患病率、发病率、病死率和死亡率的水平和消长), 研究环境因素的致病或保护作用, 对特定的群体进行的调查。

调查研究的研究者只是对研究对象进行客观的观察和记录, 而不施加任何干预因素。调查研究必须结合专业进行周密的调查设计。

调查设计是调查研究工作的先导和依据, 也是调查结果准确可靠的保证。调查设计的特点即研究因素是客观存在的, 如职业、地域、民族等; 因为不能用随机化分组平衡混杂因素对调查结果的影响, 因此重点是调查表、抽样方法的设计以及分析方法的选择。调查研究必须明确调查目的、制定合理的抽样框、科学的抽样方法和完善的调查表。通常, 需将调查研究的目的转化为拟分析的指标, 再将分析指标转化为调查项目, 并制定调查表进行资料的收集、整理和分析。完整的调查设计包括调查计划、组织计划、整理计划和分析计划, 分析计划主要涉及资料的统计分析方法。

(2) 实验研究: 研究者根据研究目的, 主动地对研究对象施加干预因素, 并控制非干预因素的影响, 以观察干预因素作用的研究。实验研究的特点是: 研究者能人为设置处理因素, 研究对象接受处理因素的种类或水平是由随机分配决定的。因此, 实验研究能够更有效地控制误差, 使多种实验因素包括在较少次数的实验之中。

广义的实验研究包括动物实验、临床试验和社区干预试验。在内科学研究中经常需要临床试验研究, 临床医学研究的内容很广泛, 包括病因学、诊断学、疗效和预后诸领域的研究。比如临床上常做疗效评价, 即疗效研究, 其内容既有药物、手术、理化因素的效应, 也有营养、护理等辅助措施与预防措施的作用, 也可以是对完整的一组治疗方案或一种特定形式的治疗措施的研究。

疗效研究的指标, 根据不同的目的可以是: 生存或死亡, 痊愈或未愈, 有效或无效, 症状或体征的存在或消失, 生理、生化指标的变化及副作用等。评价临床疗效的优劣, 应注意实验设计的类型、被试因素(药物、手术、理化因素等)的科学性、受试对象的代表性及其诊断的正确性、疗效指标的统一性和可靠性。

实验设计的要点是实验的三要素、四原则、实验设计类型等内容。受试对象、实验因素和实验效应被称为实验设计的三要素。实验研究的目的就是要阐明某种或某些实验因素对什么样的受试对象产生了怎样的实验效应。受试对象的纳入与排除标准的制定、实验因素的选定和重要非实验因素的控制、主要指标的确定主要取决于专业知识。四原则是随机、对照、重复与均衡原则。这些原则的作用就是为了保证科研工作能“透过现象看本质”。在调查或实验过程中, 严格遵循这些原则是科研成功的关键所在; 实验设计类型是由实验因素和水平决定的, 它是科学而又巧妙地安排实验研究的象征, 内科研究中常用的是单因素两水平实验设计。

1.3 内科研究方法

内科学是临床医学的一个专科, 几乎是所有其他临床医学的基础, 也有医学之母之称。因此, 内科科研的涵盖面也是非常广泛的。内科科研包括对不同系统疾病的诊断、治疗、利用现代医学的科学方法研究疾病的病因和发病机制、临床表现、诊断和鉴别诊断、治疗及预防。在内科研究中, 会涉及对呼吸系统、循环系统、消化系统、泌尿系统、血液系统、内分泌系统、代谢系统等相关课题的研究。

虽然近年来有关内科研究课题逐年增加, 但在质量上还远远赶不上整体科学发展的水平。很多为低水平、重复性的研究, 研究方向转化快, 不成规模和体系。很多课题在统计设计上存在问题, 比如缺乏合理的随机对照, 样本量过小等; 还有一些存在统计学分析方面的错误, 使得结论缺

乏可信度。有些课题从题目来看似乎非常有吸引力，但却因为设计或统计分析不正确而得不出正确的结果，浪费了人力、物力和财力。

1.3.1 选题内容

选题是进行科学研究的最重要、最有决定意义的一步，是每项科研工作的起点。它在一定程度上反映了科学研究的水平和研究成果的价值，同时也决定了最后论文的水平。选题应充分考虑和事先做好调查、阅读相关资料、了解课题背景和研究方向等。

科研选题应根据选题原则和过程，确定研究的具体问题。爱因斯坦说：“提出一个问题，往往比解决一个问题更重要。”这充分说明了科研立项的重要性。内科科研一般来自内科学实践中遇到和发现的问题，目的是解决实际问题。科研的核心是创新、不重复，创新是科研的灵魂和核心。对于前人已经做过的研究，如果不能在前人的基础上有所创新，就没必要继续研究。

内科研究根据其学科特点又可划分为基础研究和临床研究。前者一般为病因、病理、生理、生化、药理、生物、寄生虫和流行病学等实验研究，主要通过动物实验进行。后者对临床上某种疾病病例的病因、临床表现、分型、治疗方法和疗效观察等进行分析、讨论，总结经验教训，并提出新建议、新见解，以提高临床疗效，主要通过人体试验进行。但是，二者又是相辅相成的，临床研究并不排除基础研究的内容。比如在研究个体和群体对某病的易感性时，观察有关基因的多态性，采用目前尚未在临床普遍开展的分子技术，是理所当然的。再比如呼吸道合胞病毒感染(RSV)至今无有效治疗，将核酸技术移植于抗病毒领域，成功合成抗RSV特异性脱氧核酶，为治疗RSV感染性疾病提供新途径。相对来说，在临床实践中选题更加容易。可以说，临床工作是医学课题研究取之不尽的源泉，因为临床中会经常遇到许多需要解决的实际应用问题或理论问题，对此，只要从本学科实际出发，用心思考，会从中产生很多好的主题。这包括：(1)探讨发病机制与预后情况；(2)分析临床症状与表现；(3)研究诊断方法和治疗方法；(4)疾病的多因素分析等。另外，临床研究选题可以从常见病、多发病及严重危害人民健康的疾病出发。比如，近年来心脑血管疾病、恶性肿瘤、肝炎、糖尿病、老年痴呆及骨质疏松症等疾病是多发病和常见病，对人民健康危害严重，可以选择其中的一种病，深入开展防治规律、方案，手段及药物的研究。对有可能对社会造成重大影响的疾病，如艾滋病、性病及戒毒等也要予以重视，进行针对性研究。

内科研究尤其是临床内科学的研究，往往需要时间的验证，因此应该强调研究方向的相对稳定性。一个课题一旦确定以后，就应该制定一个近期、中期和远期的目标，切忌急功近利。

内科科研课题选题得当，将达到事半功倍的效果；而选题不当，则可能使研究无法实施，事倍功半。如果选择了没有意义或意义不大的课题，花再大的精力，也没有多大价值和意义；如果选择了自己不熟悉的课题，就会无从下手，“啃不动”，甚至中途夭折；如果选择了大而空的课题，就可能力不从心，效果甚微。

1. 科研选题的思路

选题是科学研究过程中的重要环节，通常用“战略性”、“关键性”、“决定性”等词句来形容这个环节。任何一项研究活动，必须提出要研究的目标。科研课题决定了科研内容的设计，指导科研工作的全过程，影响科研的成败、优劣及其价值。选准一项课题，研究人员必须了解与课题有关的历史，熟悉其现状和发展趋势，知道还存在那些有争议和未曾解决的问题。提出一项研究课题，还要对研究结果有较大的预见性；估算可能遇到那些复杂情况，准备一些如何应变的对策。科研选题的质量，实际上是研究能力和研究水平的综合表现，能较客观地反映科技人员的综合素质。由于医学科学的范围广泛，学科之间各有特点，选题要求也不尽相同。

科研选题，若按秦伯益院士的说法，则只要三句话、九个字，选题就可以成立了。第一句话是

“有人要”，第二句话是“没人搞”，第三句话是“我能搞”。第一句话指理论性研究要体现人们认识自然的需要、学科发展的需要；应用性研究要体现社会的需要，反正是有人要。第二句话指有社会需要但没人搞，包括虽然有人搞，但还没搞成或搞得不很好。第三句话指我有能力搞。而且这三句话都要具备。如果是有人要，没人搞，我也搞不了，这样的题目就等于白做；如果是有需要，我能搞，别人已经搞好了，那就是重复模仿。不过，模仿也有不同的模仿，原子弹和载人飞船的模仿就了不得；医学上若能开展多器官移植或某些高难度手术，也是有价值的模仿选题。

2. 科研选题的原则

科研选题有几个带共性的原则：创新性原则、科学性原则、需要性原则、可行性原则和合理性原则。

1) 创新性原则

没有创新就没有科研，通常人们把创新性喻为科研工作的“灵魂”、“精髓”和“难点”。目前，国家自然科学基金的申请或其他研究项目的投标，强调资助有创新性的课题。所谓创新，通俗地说，就是“人无我有，人有我优，人优我新”。“有所发明，有所发现，有所创造，有所前进”是创新性的高度概括。“有所发明”的要点是前所未有，指过去从来没有的东西，现在被发明了，如第二军医大学附属长海医院先天性心脏病封堵器的研制，研究者通过10余年的不断探索和实践，最终实现了科研成果到产品的转化，形成了具有自主知识产权的产品。改变了先心病介入治疗器械长期依赖进口产品的局面，目前，每年应用该研究成果治疗1万余例国内、外的先心病患者。“有所发现”的要点是未被认识，指过去本来就存在的客观事物，但现在才被认识到，如幽门螺杆菌的发现，虽然该菌早就存在，但一直未被人们所认识。科学家巴里·马歇尔(Barry J. Marshall)和罗宾·沃伦(J. Robin Warren)以100例接受胃镜检查及活检的胃病患者为对象进行研究，证明这种细菌的存在确实与胃炎相关。此外，他们还发现，这种细菌还存在于所有十二指肠溃疡患者、大多数胃溃疡患者和约一半胃癌患者的胃黏膜中。经过多次失败之后，1982年4月，Marshall终于从胃黏膜活检样本中成功培养和分离出了这种细菌。为了进一步证实这种细菌就是导致胃炎的罪魁祸首，Marshall和另一位医生Morris不惜喝下含有这种细菌的培养液，结果大病一场。二人的创新意识和为人类医学事业贡献的精神，感动了医学界。二人也因为其创新而获得2005年的诺贝尔生理学 and 医学奖。“有所创造”、“有所前进”的要点是在原有的基础上，有新的进步、新的发展。基础研究方面的创新性是指对学科发展有价值的理论性问题；应用基础研究的创新性表现在能够指导实践的，能提高预防、诊断、治疗、康复效果的新技术、新方法、新材料、新器材、新见解、新理论或新手术方案等。这些都是“有所前进”的表现。没有新意的、简单重复的课题不能算是科学研究。应当说明，有些课题虽有重复，但仍有意义和价值。例如，一些难度大、技术要求高、能够标志技术发展阶段，或在国内，或在本地地区必须填补空白的项目；一些新近提出的、尚未充分成熟、有待完善和验证的成果等。

但是，如何进行创新还没有固定的成规可循。其实，“创新”未必是惊天动地的大发现、大发明，创新并不是“高不可攀”，创新是有不同层次、不同阶段、不断升华的过程，在医学科研领域创新永无止境。

2) 如何才能创新

创新的灵活性很大，首先，要全面地掌握信息学科领域的信息；其次，要有高的起点。“新”与“旧”是相对而言的，是相互比较中存在的。从主观愿望出发，研究者并不希望自己所选的是重复别人的课题，其所以会出现重复他人的结局，主要是不了解有关的动态和信息，以致辛辛苦苦，自以为是创新性的结果，事实上是别人早已发表过并为许多实践证明过的东西。要了解有关科技信

息，当然要阅读有关文献资料，在文献浩如烟海的今天，要学会文献检索和使用的方法。借鉴文献资料可以学得经验，吸取教训，少走弯路，避免重复，提高研究工作的起点。除文献查阅以外，应重视学术性交流活动。在国内许多专业性和专题性医学学术会议上反映的信息比较集中，这些信息一般比公开发行的期刊早一年左右。此外，有针对性地和一些专题研究者请教，往往可以避免明显的选题重复。

由于科研工作有一个较长的周期，尽管在选题时尚无有关报道，但很可能在科研过程中或结束时，别人公开发表了类似的科研成果，难以预料地落入重复别人的结局。因此，选题时若能注意“起点要高”，知道同一研究领域中，哪些问题尚未解决，选择从最前沿出发的研究课题，可以减免上述不愉快的结局。即使有了一定的重复，由于课题的起点高、学术价值大，处在学术发展阶段性的前沿，正需要展开讨论和验证，这些研究结果不致沦为低级重复的下场。

要创新就要发挥分析、判断、筛选等综合思考作用，下列几点可供参考。

对于一些需要性很强但创新性较难的课题，比如医学研究中的肿瘤问题、衰老问题、免疫排斥问题、心血管问题、糖尿病问题等“老、大、难”的课题，应特别重视文献信息的熟悉，结合内科学和个人的条件，从中选择可以进行的、尚有创新意义的分题，避免重复。对于一些不为人们注意的问题，创新性较易，尤其要论证是否具有“需要性”原则，反对单纯猎奇、忽视实际价值的倾向。如果花了大量人力物力得到的新资料或新方案，在临床应用上毫无意义，提出的新见解、新理论又不能在学科发展中起作用，那么这种研究是没有价值的，是选题中力求避免的大忌。

在前人的基础上，发展相对性创新，也是创新的思路之一。科学技术上的创造发明，绝大多数是在前人已经建立的基础上进一步发展的。一个尚处在初级阶段的研究人员，在阅读文献或听取别人经验时，总觉得所有的论文和经验都很完善，感到自己很难有所改进和创新。一个日趋成熟的科研人员，已有较丰富临床实践经验、具有独立思考 and 独立工作能力的内科医生，应认识到学术发展是没有止境的，文献资料只能说明阶段性的成就，绝不可能是十全十美的，应在继承中创新、发展。在积累了经验和教训、提高了分析判断能力的基础上，研究人员往往可以发现文献或参考书、教科书的缺陷或错误，敢于补充、改进、修正前人的结论。若将这些认识和设想，通过科研设计和科研实践加以证实，就是一种创新性研究过程。

借鉴、移植为创新打开了新的思路。借鉴或移植也有创新因素，比如工业用的超声探伤借鉴到医学领域，有了医学的B超检查；早期的胃镜、气管镜仅用于诊断，由于借鉴和移植性的创新，现在已经广泛地应用于介入和微创外科治疗。

在前人已经研究过的基础上，启用新技术和新仪器，借助于新的技术、方法或仪器，可以得到更加科学的新资料，使研究工作得到深化和发展。在分子生物学高新技术和新型医疗仪器设备发展很快的时代，若运用得当可以找到大量有新意的课题，比如采用不同性能的显微镜、不同的检测手段、不同的新试剂等都是创新。

“边缘地带”、“交叉地区”，常有大量的“空白点”。在这个地带容易找到有创新性的选题。例如我校获得国家科技进步二等奖的“显微外科解剖学研究”，所观察的范围是大体解剖学用肉眼观察已有困难，而组织学用低倍镜还有余，不屑一顾的地域；但是这个领域却是显微外科的急需应用，又缺乏可查阅的资料。因此，只要显微外科临床有所需要的，都可以做，做出来以后都是既能“填补空白”，又有实用价值的成果。交叉学科更有广阔的天地。例如，将探测工业材料断裂损伤的超声技术或核磁共振成果用于人体的影像诊断；将研究工程材料力学的方法用于生物材料的测试；按照内窥镜“管中窥豹，只见一斑”的特点，开展微创外科解剖学的研究，升华为“窥斑知豹，观微知著”，既见树林，又识森林；将计算机技术与人体结构相结合，开展了“数字人的研究新领域”。在内科学领域，既要注意“基础与临床结合”，又要注意“中西医结合”，同时还要将“老学科”

与“新技术结合”，结合就会有大量可供开发的新课题，所谓结合，要特别重视“边缘地带”和“交叉地区”的结合。

最后，还要注意借鉴前人的思路，开拓自己的思路，以达到创新的目的。“他山之石，可以攻玉”。很多事物可以举一反三，触类旁通。带着创新的欲望和目的，努力学习前人报道过的实验手段、设计方案、动物模型、技术方法、影响条件、监测指标、评价标准等，均可以启迪自己的思路。取其所长，补其所短，结合自己的特点和优势，融会贯通后加以改进，也有广阔的创新天地。作为科技工作者，首先要汲取前人累积的经验和教训。但是，更重要的是有创新性思维、有解决问题的能力。

3) 科学性原则

科学性原则是选题中最根本的原则。如果学术论文的科学性有问题，通常在科研成果鉴定时，称为“一票否决制”。科研是探索未知，但必须符合客观实际，反映客观规律，要有充分的依据，可以用科学研究的方法去解决，而且经得起重复和实践检验。当前，虚假浮躁风已在侵蚀学术界，特别要反对伪科学和弄虚作假的行为。

科学研究有别于科学幻想和文学艺术，一切“长生不老药物”、“绿豆能治百病”等耸人听闻的“课题”均缺乏科学性，是不能成立的；现代医学史上出现过的“活质学说”、“经络小体”、“鸡血疗法”、“特异功能”等轰动一时事物，均因科学性不足而失败，应引以为戒；“睡眠疗法”、“卤碱疗法”、“针刺疗法”、“气功疗法”都要有明确的范围，过分夸大就不符合科学性。

4) 需要性原则

选题应有明确的需要性，因为医学有重要的“救死扶伤，治病救人”任务。结合我国目前以经济建设为中心的阶段，在重视基础理论研究的同时，还要强调科研的社会效益和经济效益，应努力促进将科研成果转化为生产力。虽然有些基础研究课题，在开始时可能根本没有或没有考虑任何实用价值，甚至学术价值也不很明显，但在日后却显示出重大的学术意义和巨大的应用价值。例如居里夫人刚开始研究镭的放射性能时，并没有考虑到在日后医学临床上有那么广泛和重要的用途。

需要性原则就是科研的目的性，明确为什么要进行这个研究项目。从广义上说，是为了人类生存、社会建设或科学发展的需要。作为医学科研，重点要解决医学基础理论、疾病预防、诊断、治疗和康复等有关问题。

医学科研中，临床诊治的实际需求就是医学研究选题最有活力的源泉。在临床诊治工作中，只要精益求精、尽力探究、穷追不舍，会发现当前医学无法解答的问题层出不穷，有许多是学者和文献未能认识或记载的问题。对于病人身上的疾苦，确实有许多当前医学尚难解决的问题。这些问题促使学者们寻求新的、更有效的诊治方法，若能结合这些问题进行选题，研究目的都有很明确的针对性，都属于需要性很强的题目。

但是密切结合临床需要，不等于向任何实践经验丰富的医师请教即能达到目的。“同阅一卷书，各自领其奥；同作一题文，各自擅其妙”。我们也都当过病人，见过不少“医师”，也知道“这个医师”与“那个医师”之间也有很多差异。为数不少的医疗实践者，由于工作条件，习惯于传统的诊治方案，满足于现有技术水平，不准备进行改进或创新，他们就提不出科研选题。不满足于现状、勇于探索、勤于思考、迫切要求提高诊治水平的医师们，才能提供需要性很强的研究课题。这些医学工作者，较集中地出现在各种学术会议场合。为此，适当参加或列席学术会议，多与开拓精神很强的学者们交换意见，也能增加选好课题的良好机会。

5) 可行性原则

选题要从实际出发，脚踏实地，结合本人和本单位情况考虑有无实现课题研究的主客观条件。

在主观条件上,要有自知之明,包括本人过去的工作基础、学术水平、技术能力,要量力而行。客观条件上要考虑经费、设备、资料、指导能力、后勤保障和可协作条件等。科研周期还有很强的时间性,必须按规定的期限完成,不得拖延。为此,从事动物实验的,要考虑有无合格的动物供应、饲养条件等;从事高技术新技术实验的,要考虑试剂供应、仪器设备条件等;从事临床试验研究的,要考虑有无充足的病例来源、有无必要的检测条件等;从事调查研究的,要考虑有无有价值的对象和处理数据的条件等。现代科研工作发展很快,科研手段日趋综合,任何单位都不可能应有尽有,包罗万象,许多仪器设备和专门技术,通过有可能争取到的协作途径加以解决。

6)合理性原则

合理性原则也称为效益性原则,是一种讲求效益的原则。日常生活知识告诉我们,并不是对人类最重要的东西,就是最值钱的东西,例如“空气”和“水”,是生命攸关的东西,但目前还卖不出大价钱,原因是“物以稀为贵”。一切预期的科研成果,必须在理论意义、社会效益和经济效益上具有价值。按市场经济规律估价,一项科研成果能在各个方面都有优异效益固然很好,但这种十全十美的成果所占比例不可能很大。因此,选题立论时,起码要求研究得出的成果,在某一方面或某一点上优于传统的东西。例如,在研究一种新技术治疗糖尿病是否优于旧技术时,应要求新术式比旧术式在功能好、疗程短、费用省、操作简单、安全可行、病人痛苦少等方面,有多项或单项的优越性。如果预期的研究结果在各方面均今不如昔,那么这个选题就不符合合理性原则,这种创新就缺乏实际效益,不能成立。

上述选题的基本原则,是内科学选题时应当注意的事项。由于各个科室研究工作基础和发展要求不尽相同,选题要求是有所区别的。但总的来说,大部分带共性的必须遵循的原则,是内科学研究选题可以参考的基本知识。

1.3.2 研究方法及统计学应用

在我国,内科学研究起步较晚,从目前对内科学研究的回顾总结中可以看到,内科学研究以临床试验性的研究为多。

1. 统计设计

在内科学科论文中,以住院病人为研究对象的试验性研究很多。这些研究在统计研究设计方面还存在很多问题。主要有以下几个方面。

(1)随机分组方法不合理。在临床试验研究中,以住院病人为研究对象,在很多论文中,作者仅仅提到“将受试者随机分为两组”,并没有阐述具体的随机分组方法;或以某一段时间的病人为对照组,另一段时间的病人为试验组,还有的研究者按照患者入院顺序对受试对象进行分组,这些都是错误的。一个实验一旦违背了随机原则,那么对照与均衡原则必然会受到人们的质疑。

(2)没有进行合理的样本含量估计。正确估计样本含量体现了研究设计严格遵守了重复原则,可以降低研究中的抽样误差,同时足够的样本量也是实验研究中保证组间均衡性和结果的重现性的基础。若样本量过小,观察指标的平均值则不稳定,意味着抽样误差大,推论总体的精密性与准确性都比较差,统计检验的效能(即发现客观上存在的差别的能力)低,实际存在的差别不易真实地显露出来;样本含量过大,会增加实际工作的困难,浪费人力、物力、财力和时间,虽然减少了抽样误差,但由于过分追求数量,可能引入更多的混杂因素或因工作粗枝大叶导致科研资料不准确,对研究结果造成不良影响。

(3)实验设计类型单一。在内科学科论文中,单因素实验设计比较常见。当存在重要的

非实验因素时,研究者可以考虑使用随机区组设计、拉丁方设计、含区组因素的析因设计等设计类型来合理安排实验因素和控制重要非实验因素。

2. 资料的表达与描述

资料按性质分为定量资料和定性资料两大类。研究者多以统计表的形式表达整理资料,定量资料以平均数和标准差来描述资料的平均水平和变异程度;定性资料常以列联表的形式呈现,其内通常既有绝对数又有相对数。统计图也是表达资料的常用方法,但在内科学论文中却被应用得相对较少。

存在的问题有:

(1)研究者在使用平均数和标准差时,并没有考虑资料是否服从正态分布,对于呈偏态分布的资料同样采用了该指标。

(2)计算相对数时,分母过小。例如,用某药治疗某病患者,5例中有3例治愈,计算治愈率为 $3/5 \times 100\% = 60.0\%$;如果有4例治愈,则其治愈率为80%。在分母例数很少的情况下,个别的偶然因素会导致结果的变化,只有分母例数较大时,计算的相对数才比较稳定。在例数较少时,通常直接用绝对数表示。如果要用相对数表示,则需列出其总体率的置信区间。

(3)统计表不符合规范。主要表现是横、纵标目设置不合理。

(4)适宜用统计图表达的资料仍然用统计表表达。统计图较统计表更加直观,能够给读者留下更深刻的印象,使人一目了然。

(5)涉及多个因素的定性资料,研究者没有采用高维列联表表达资料。

3. 统计分析方法的合理选择

t 检验和卡方检验是内科科研工作者使用最多的两种检验方法。秩和检验、方差分析、相关与回归分析也有一定的应用。个别论文中还使用了多重回归分析和某些多元统计分析方法。

存在问题有:

(1)滥用 t 检验。凡是定量资料,绝大多数研究者采用的是 t 检验,但很少有研究者考虑 t 检验的前提条件,即独立性、正态性和方差齐性。不仅如此,还应注意资料所取自的实验设计类型是否为单组设计、配对设计和成组设计;也有研究者用 t 检验处理定性资料;当实验中涉及时间这样的重复测量因素时,研究者仍然使用 t 检验分别比较各时间点上的数据,这些都是错误的做法。

(2)滥用卡方检验。对于总例数小于40或存在理论频数小于1的四格表资料,1/5以上格子内理论频数小于5或存在理论频数小于1的 $R \times C$ 列联表资料应使用Fisher精确概率法检验,但研究者仍然使用卡方检验。结果变量为多值有序资料适宜用秩和检验进行统计分析,也有研究者采用卡方检验。

(3)不用统计分析方法。有的研究者用简单的百分比直接进行比较,就得出结论。

(4)在调查研究中,涉及多个因素时,研究者采用单因素分析方法,割裂了资料的整体性,忽略因素间的交互作用,得到的结论令人怀疑。

(5)论文中还存在统计量计算错误的情况,或不提供统计量的值,仅呈现 P 大于0.05或小于0.05。

统计学是内科科研结论是否准确可靠的重要保证。然而,我国内科科研中统计学的应用状况却令人担忧。主要原因在于内科研究工作者没有系统地学习医学统计学的基础知识,不能合理运用实验设计与统计分析方法。因此,广大内科研究工作者急需学习统计学知识,逐渐形成与统计工作者合作的意识,科学合理地进行科研工作,避免盲目实验造成的人力、物力、财力和时间的浪费,避免学术造假。

1.4 如何学习统计学

1.4.1 统计学的整体构架

学习统计学时，若对其整体架构一无所知，就好比一个站在迷宫外面的人，不知其内暗藏着哪些陷阱和玄机。下面用图 1-2 来掀开迷宫的顶盖，使迷宫暴露于光天化日之下。

仅从统计学本身来说，其主要内容包括八大功能模块，即统计思想、三型理论、科研设计、质量控制、表达描述、统计分析、结果解释、结论陈述。每一模块知识的缺乏或掌握得不准确、不全面，都会直接影响统计学的应用质量。在这八大功能模块的每个模块之下，又分别嵌套着数目不等的子模块。所有子模块功能的详细介绍和解说，就是统计学的全部内容。

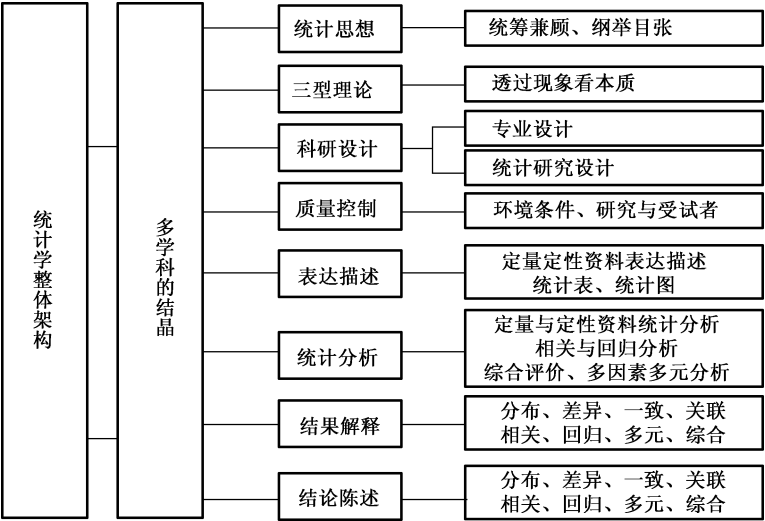


图 1-2 统计学整体架构图

1.4.2 正确的统计思想

为什么正确运用统计学，人们就能够透过事物现象看清其本质？这是因为统计学中蕴含了丰富的辩证唯物主义思想。例如，在拟进行一项科研课题的研究时，从统计学角度来看，首先就要把一切可能会影响结果的因素都梳理出来，并认真加以排查，淘汰掉那些肯定没有多大作用的因素，将剩下的因素同时纳入研究，并对其进行科学合理的安排。这实际上反映了统计学在处理问题时，注重把握事物的“系统性和代表性”特征；而被研究对象往往是无限多的，实际研究时只能考察其中的很小一部分，统计学要求将被研究对象按照某些重要特征（如地区、性别、年龄等）进行分层，再从各层中随机抽取一定数量的个体构成研究对象的全体，这实际上反映出统计学巧妙地利用了“随机性和均衡性”，以便揭示隐藏在事物内部的本质特征。由部分具有代表性的个体所提供的信息，借助统计学方法去推论总体的规律性，这一特性反映统计学具有“延展性”。统计学期望弱化静态思维，强化动态思维；突破正向思维，巧用逆向思维；跳出简单思维，步入复杂思维；活用横向思维，发展纵向思维。诸如此类的思想和做法，都是为了达到“透过现象看本质”的目的，可概括为统计思想，而不是单纯的“统计计算”。

具体地说，统计思想包含两个大的方面：其一，辩证法与多学科的结晶；其二，三型理论。统

计学中反复强调的若干原则,就是为了保证研究结果能真正反映客观实际,即透过事物的现象看清其本质,这正是唯物辩证法的核心。所谓多学科结晶,是指统计学中不仅包含数学味很浓的“概率论与数理统计”,而且,包含唯物辩证法、基本常识、各科专业知识、计算机科学和信息学等。也就是说,要想真正学好统计学、用对统计学,应尽可能多掌握一些除数学和统计学之外的其他学科知识,特别是会不自觉地正确运用基本常识。因为许多科研工作者在科研工作中,如其说他们经常会犯统计学错误,倒不如说他们最易犯的是基本常识错误。例如,某临床医生为了研究某利尿药的疗效,在进行临床试验设计时,仅观测服试验药与对照药的受试者排尿量的多少,但不观测试验组与对照组中患者每天的饮水量。

1.4.3 三型理论

所谓三型理论,实际上就是一个正常人看事物的一种思维方法,即任何事物和现象都由“表现型、原型和标准型”组成,只要研究者弄清了与每个具体问题对应的这三型分别是什么,再有的放矢地去解决它,一般不容易出错。将三型理论巧妙地运用于统计学中,可以很容易达到“化繁为简、由表及里和透过现象看本质”的功效。例如,有一位70多岁的教授,他右边头发绝大部分是黑色的,而他头部其他部分的头发几乎全白了。这是什么原因呢?这显然是一个表现型,而原型是什么呢?原型应当是,整个头部都应该几乎是白发,现在右头部出现了一些黑发,是由于他在思考问题时习惯用右手由前向后梳理头发所致。其标准型就是因为经常梳理右边头发时,使右边头部皮肤得到运动,使其下部的肌肉和血管经常保持很好的柔韧性,达到“疏通经络、促进新陈代谢”之功效。换句话说,只要人们坚持每天用手指尖适度用力梳理头发,可以保持头发不会过早变白。当然,这个个案是否具有普遍性,有待基于正确而又完善的科研设计指导下的大样本试验来验证。

1.4.4 学好统计学的要领

单从应用角度来看,学好统计学并非很难的一件事。关键的要领是始终做一个“正常人”。第一,很多统计学理论和方法(特别是科研设计)是建立在“基本常识”和“基本专业知识”基础之上的。第二,要消除恐惧心理。因为复杂的统计计算可借助现代统计软件来实现,只需把学习的注意力和精力花在常用统计学概念和方法合理选择上。第三,掀开“迷宫的顶盖”,让其一切暗藏着的“机关”暴露无遗,即了解统计学的整体架构,使其重点突出、层次分明。第四,学习时应注意循序渐进、由浅入深,要善于实践,模仿书中的例题,举一反三,反复实践。第五,应在正确的统计思想和三型理论指导下,针对具体问题(包括设计和资料的背景及所具备的前提条件)和分析目的,构建最优策略解决所面临的实际问题。

(王素珍 曹文君)

参 考 文 献

- [1] 陆再英,钟南山. 内科学(第7版)[M]. 北京:人民卫生出版社,2008:121-123.
- [2] 王汝芬,陈景武,李建国等. 医药科学研究统计方法[M]. 山东:山东科学技术出版社.1994:1-17.
- [3] 杨锡强. 我国儿科临床医学研究的现状与思考[J]. 中华儿科杂志,2008,46(5):321-4.
- [4] 胡良平主编. 检验医学科研设计与统计分析[M]. 北京:人民军医出版社,2004:321-4.
- [5] 胡良平主编. 口腔医学科研设计与统计分析[M]. 北京:人民军医出版社,2007:321-4.
- [6] 胡良平主编. 科研设计与统计分析[M]. 北京:军事医学科学出版社,2012:3-10.

第 2 章 常用内科统计设计与数据分析

在内科科学研究中，要合理应用统计学方法对资料进行处理，需要结合试验目的，认真考察资料是否满足参数检验方法的前提条件(独立性、正态性和方差齐性)，并需正确识别统计资料的类型，正确辨析资料所对应的试验设计类型。本章将结合一些实例对内科研究中常用的一些科研设计和数据分析方法做一简单介绍。

2.1 单因素 K 水平 ($K \geq 2$) 设计与分析

2.1.1 应用目的

单因素 K 水平 ($K \geq 2$) 设计是一种将研究对象分配到某试验因素各水平组中去的一种设计方法。单因素 K 水平 ($K \geq 2$) 设计只涉及一个试验因素，未对其他任何重要的非试验因素进行有计划的安排，仅希望通过随机化抽样和/或分组来平衡所有非试验因素在各试验组之间对观测结果的影响和干扰。

2.1.2 设计方法

根据所考察的一个试验因素与受试对象是否直接有关，可细分为两种情形：(1) 试验因素独立于受试对象(如试验因素为药物种类或剂量大小)，从基于研究目的而定义的具有同质性的总体中完全随机抽取受试对象并再将它们完全随机分配到各小组中去；(2) 试验因素非独立于受试对象(如试验因素分别为疾病严重程度或血型或性别或职业)，需从该试验因素各水平定义的各子总体中随机抽取受试对象，它们自然形成了各试验组中的受试对象，不需要接受第 2 次完全随机分组。

2.1.3 具体设计过程

- (1) 将全部研究对象按某种特征(如动物的体重、患者的就诊顺序等)编号。
- (2) 从随机数字表或利用计算机、计算器产生伪随机数。
- (3) 将随机数抄录在需要被分组的受试对象的编号之下。
- (4) 按事先规定的随机分组规则将各受试对象标注出应分入的组别。

【例 2-1】 将 15 例糖尿病患者随机分到 A、B、C 三组。其具体操作过程如下。

- ①按糖尿病患者的空腹血糖水平编号，分别为 1,2,3, ...,14, 15。
- ②从随机数字表中的任何一行、任何一列、任何一个数字开始随机取 15 个 2 位数的随机数，若遇到相同的数则舍弃后重新取随机数。如从第 5 行第 11 列开始向右依次取 15 个 2 位数的随机数，分别为 73,37,32,04,05,69,30, 16,09,05(舍弃),88,69(舍弃),58,28,99, 35,07。
- ③将随机数从小到大排列后得到序号 R ，见表 2-1。并事先规定 $R = 1 \sim 5$ 的患者分到 A 组， $R = 6 \sim 10$ 的患者分到 B 组， $R = 11 \sim 15$ 的患者分到 C 组。

表 2-1 15 例糖尿病患者的完全随机分组结果

患者编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
随机数	73	37	32	04	05	69	30	16	09	88	58	28	99	35	07
序号 R	13	10	8	1	2	12	7	5	4	14	11	6	15	9	3
处理组别	C	B	B	A	A	C	B	A	A	C	C	B	C	B	A

2.1.4 统计分析方法

对于单因素 K 水平 ($K \geq 2$) 设计一元定量资料, 当各组数据来自于正态总体且各总体方差相等时, 可采用单因素 K 水平 ($K \geq 3$) 设计一元定量资料方差分析 (analysis of variance, ANOVA) 或者单因素两水平设计一元定量资料 t 检验; 对试验效应指标有影响的基线指标, 可将其作为协变量, 采用单因素 K 水平 ($K \geq 2$) 设计一元定量资料协方差分析 (analysis of covariance) 进行比较。当资料不满足正态性和方差齐性时, 可将原始数据进行转换; 当转换后的资料满足参数检验的前提条件时, 再对其进行单因素 K 水平 ($K \geq 2$) 设计一元定量资料方差分析, 或直接采用单因素 K 水平 ($K \geq 2$) 设计一元定量资料秩和检验 (Rank sum test)。当组数 = 2 时, 用 Wilcoxon 秩和检验; 当组数 ≥ 3 时, 用 Kruskal-Wallis H 秩和检验。

2.1.5 SAS 实现程序及结果解释

1. 实例分析及结果解释

【例 2-2】 对 10 例肺癌患者和 12 例矽肺 0 期工人用 X 光片测量肺门横径右侧距 R_1 值 (cm), 结果见表 2-2。问肺癌患者的 R_1 值是否高于矽肺 0 期工人的 R_1 值?

表 2-2 肺癌患者和矽肺 0 期工人的 R_1 值比较

编 号	1	2	3	4	5	6	7	8	9	10	11	12
肺癌患者	2.78	3.23	4.20	4.87	5.12	6.21	7.18	8.05	8.56	9.60	·	·
矽肺 0 期工人	3.23	3.50	4.04	4.15	4.28	4.34	4.47	4.64	4.75	4.82	4.95	5.10

假定此定量资料不满足参数检验的前提条件, 则所需要的 SAS 程序如下:

```
data nkkysj2_2;
  input x c $ @@;
datalines;
2.78 A 3.23 A 4.20 A 4.87 A 5.12 A 6.21 A 7.18 A 8.05 A 8.56 A 9.60 A
3.23 B 3.50 B 4.04 B 4.15 B 4.28 B 4.34 B 4.47 B 4.64 B 4.75 B 4.82 B
4.95 B 5.10 B
;
run;
proc npar1way wilcoxon;
  var x;
  class c;
run;
```

SAS 运行结果及结果说明:

The NPAR1WAY Procedure

(第一部分)

Wilcoxon Scores (Rank Sums) for Variable x
Classified by Variable c

Sum of		Expected	Std Dev	Mean	
c	N	Scores	Under H0	Under H0	Score
A	10	141.50	115.0	15.161469	14.150000
B	12	111.50	138.0	15.161469	9.291667

Average scores were used for ties.

(第二部分)

Wilcoxon Two - Sample Test

Statistic 141.5000

Normal Approximation

Z 1.7149

One - Sided Pr > Z 0.0432

Two - Sided Pr > |Z| 0.0864

t Approximation

One - Sided Pr > Z 0.0505

Two - Sided Pr > |Z| 0.1011

Z includes a continuity correction of 0.5.

(第三部分)

Kruskal-Wallis Test

Chi-Square 3.0550

DF 1

Pr > Chi-Square 0.0805

结果说明：整个结果可分为三部分。

第一部分给出了原始变量值进行秩转换后的秩得分值的一些简单描述性统计量。

第二部分为 Wilcoxon 两样本检验正态近似和 t 近似的结果，涉及的统计量有较小样本的秩和 (Statistic)、正态近似法检验的 Z 值和检验统计量所对应的 P 值 (Two - Sided Pr > |Z|)。本例的 Statistic 为 141.500, $Z = 1.7149$, (正态近似) $P = 0.0864 > 0.05$; (t 近似) $P = 0.1011 > 0.05$; 所以不能认为肺癌患者的 R_i 值与矽肺 0 期工人的 R_i 值的差别有统计学意义。

第三部分为 Kruskal-Wallis 检验结果，这里，实际上采用了卡方近似法，该检验对两组样本和多组样本都适用，本例 $\chi^2 = 3.0550$, $P = 0.0805$, 结论同第二部分，从略。

【例 2-3】 根据表 2-3 的资料，请比较以下三种不同人群的血浆总皮质醇测定值间有无差别？

表 2-3 三种人群的血浆总皮质醇测定值 (10 ² μmol/L)		
正 常 人	单纯性肥胖	皮质醇增多症
0.11	0.17	2.70
0.52	0.33	2.81
0.61	0.55	2.92
0.69	0.66	3.59
0.77	0.86	3.86
0.86	1.13	4.08
1.02	1.38	4.30
1.08	1.63	4.30
1.27	2.04	5.96
1.92	3.75	6.62

【实例辨析】 本例为从各子总体内随机抽样的单因素三水平设计一元定量资料三个平均值或平均秩之间的比较问题。如果满足参数检验前提条件，则可采用单因素三水平设计一元定量资料方差分析对其进行处理；如果不满足参数检验条件，则可采用单因素三水平设计一元定量资料 Kruskal-Wallis 秩和检验方法进行处理。

若该资料满足参数检验的前提条件，则单因素三水平设计一元定量资料方差分析的 SAS 程序如下：

```
data nkkysj2_3_1;
  do c=1 to 3;
    input x @@;
    output;
  end;
datalines;
0.11 0.17 2.70
0.52 0.33 2.81
0.61 0.55 2.92
0.69 0.66 3.59
0.77 0.86 3.86
0.86 1.13 4.08
1.02 1.38 4.30
1.08 1.63 4.30
1.27 2.04 5.96
1.92 3.75 6.62
;
run;
ods html;
proc anova;
  class c;
  model x=c;
  means c;
  means c/snk;
run;
ods html close;
```

SAS 运行结果及结果说明：

The ANOVA Procedure					
Source	DF	Anova SS	Mean Square	F Value	Pr > F
c	2	62.54054000	31.27027000	30.80	<.0001

此结果表明，人群的血浆总皮质醇测定值间的差别有统计学意义， $F = 30.80$ ， $P < 0.0001$ 。

Level of c	N	x	
		Mean	Std Dev
1	10	0.88500000	0.4872884
2	10	1.25000000	1.05687169
3	10	4.11400000	1.30054006

此结果表明，正常人的血浆总皮质醇测定值最低，皮质醇增高症者的血浆总皮质醇测定值最

高。但它们之间两两比较结果如何，还需进一步进行假设检验。下面是采用 q 检验(即 SNK 检验)给出的结果：

Means with the same letter are not significantly different.			
SNK Grouping	Mean	N	c
A	4.1140	10	3
B	1.2500	10	2
B	0.8850	10	1

由 SNK 检验所得结果可知，说明正常人组与单纯性肥胖组的血浆总皮质醇测定值之间差别无统计学意义，这两组人群的血浆总皮质醇测定值低于皮质醇增高症者的血浆总皮质醇测定值。

若该资料不满足参数检验前提条件，则 Kruskal-Wallis 秩和检验 SAS 程序如下：

```
data nkkysj2_3_2;
input x c @@;
datalines;
0.11 1 0.17 2 2.70 3
0.52 1 0.33 2 2.81 3
0.61 1 0.55 2 2.92 3
0.69 1 0.66 2 3.59 3
0.77 1 0.86 2 3.86 3
0.86 1 1.13 2 4.08 3
1.02 1 1.38 2 4.30 3
1.08 1 1.63 2 4.30 3
1.27 1 2.04 2 5.96 3
1.92 1 3.75 2 6.62 3
;
run;
ods html;
proc npar1way wilcoxon;
var x;
class c;
run;
ods html close;
```

SAS 运行结果及结果说明：

The NPARIWAY Procedure					
Wilcoxon Scores (Rank Sums) for Variable x					
Classified by Variable c					
Sum of		Expected	Std Dev	Mean	
c	N	Scores	Under H0	Under H0	Score
1	10	96.50	155.0	22.725245	9.650
2	10	117.50	155.0	22.725245	11.750
3	10	251.00	155.0	22.725245	25.100
Average scores were used for ties.					
Kruskal-Wallis Test					
Chi-Square				18.1300	
DF				2	
Pr > Chi-Square				0.0001	

结果说明：

与两组比较相比，多组比较也会输出一些简单的描述性统计量以及 Kruskal-Wallis 秩和检验结果，但没有正态和 t 近似法的检验结果。

本例 Kruskal-Wallis 秩和检验结果为 $\chi^2 = 18.1300$ ， $P = 0.0077 < 0.05$ ，三种不同人群的血浆总皮质醇测定值的平均秩之间的差别有统计学意义，欲知详情，还需要基于平均秩的两两比较，此处从略。

2.1.6 实例的辨析与释疑

【例 2-4】 原文题目：影响慢性咳嗽患者生活质量的相关因素分析。方法按照慢性咳嗽病因诊断流程入选患者。通过支气管乙酰甲胆碱激发试验测定患者气道反应性，记录患者莱斯特咳嗽问卷 (LCQ) 评分。将 100 例慢性咳嗽患者纳入研究，比较分析不同病因导致的慢性咳嗽患者 LCQ 总分及各项分值，结果见表 2-4。试分别分析 LCQ 总分、生理领域评分、心理领域评分、社会领域评分在不同病因间的差异有无统计学意义。试辨析资料的设计类型并合理选择统计分析方法。

表 2-4 不同病因导致的慢性咳嗽患者 LCQ 总分及各项分值 (分, $\bar{x} \pm s$)

病 因	例 数	LCQ 总分	生理领域评分	心理领域评分	社会领域评分
CVA	32	11.190 \pm 3.448	3.950 \pm 0.985	3.440 \pm 1.338	3.810 \pm 1.434
EB	16	13.280 \pm 2.592	4.490 \pm 0.876	4.190 \pm 1.041	4.600 \pm 1.062
GERC	2	12.430 \pm 2.450	4.630 \pm 0.707	3.430 \pm 0.202	4.380 \pm 1.945
UACS	7	12.730 \pm 3.673	4.430 \pm 1.074	3.660 \pm 1.550	4.650 \pm 1.443
AC	14	13.180 \pm 2.664	4.480 \pm 0.320	4.250 \pm 1.154	4.450 \pm 1.478
PIC	10	13.000 \pm 2.950	4.700 \pm 1.116	3.950 \pm 1.094	4.350 \pm 1.160
其他原因	19	11.600 \pm 2.730	4.200 \pm 1.082	3.380 \pm 1.114	4.020 \pm 0.942

注：LCQ：莱斯特咳嗽问卷；CVA：咳嗽变异性哮喘；EB：嗜酸粒细胞性支气管炎；GERC：胃食管反流性咳嗽；UACS：上气道咳嗽综合征；AC：变应性咳嗽；PIC：感染后咳嗽。

【实例辨析】 本例资料的观测结果有四个指标，分别是 LCQ 总分、生理领域评分、心理领域评分及社会领域评分，均为定量资料。只有一个试验因素 (病因)，有 7 个水平，因而属于单因素多水平设计定量资料。若满足参数检验条件，则可选用单因素多水平设计定量资料方差分析进行处理；如果资料不满足参数检验条件，可进行适当的变量转换或直接选用 Kruskal-Wallis 秩和检验。

【例 2-5】 为研究某降糖药物的有效性及其合用盐酸二甲双胍片的有效性，某研究者收治 90 名 II 型糖尿病患者，并采用随机对照试验，将研究对象随机分为 3 个治疗组：第一组为该降糖药组，第二组为盐酸二甲双胍片组，第三组为该降糖药 + 盐酸二甲双胍片组。每组 30 名患者，治疗 3 个月，主要有效指标为糖化血红蛋白。测得每个患者入组前 (X) 和 3 个月后 (Y) 糖化血红蛋白含量 (%)，见表 2-5，试分析三种药物降糖化血红蛋白的效果是否不同。请问：如何合理选择统计分析方法？若有可能，用统计软件实现统计分析。

表 2-5 三组患者治疗前后的糖化血红蛋白含量 (%)

编 号	X_1	Y_1	X_2	Y_2	X_3	Y_3
	第一组		第二组		第三组	
1	10.8	9.4	10.4	9.2	9.8	7.6
2	11.6	9.7	9.7	9.1	11.2	7.9
3	10.6	8.7	9.9	8.9	10.7	9.0
4	9.0	7.2	9.8	8.6	9.6	7.8
5	11.2	10.0	11.1	9.9	10.1	8.5
6	9.9	8.5	8.2	7.1	9.8	7.5
7	10.6	5.3	8.8	7.8	10.1	8.3
8	10.4	8.1	10.0	7.9	10.3	8.2

续表						
编 号	X_1	Y_1	X_2	Y_2	X_3	Y_3
	第一组		第二组		第三组	
9	9.6	8.5	9.0	8.0	11.0	8.4
10	10.5	9.1	9.4	9.0	10.5	8.1
11	10.6	9.2	8.9	7.9	9.2	7.0
12	9.9	8.4	10.3	8.9	10.1	7.7
13	9.5	7.6	9.3	8.9	10.4	8.0
14	9.7	7.9	9.2	8.1	10.0	6.6
15	10.7	8.8	10.9	10.2	8.4	6.1
16	9.2	7.4	9.2	8.5	10.1	8.1
17	10.5	8.6	9.2	9.0	9.3	7.8
18	11.0	9.2	10.4	8.9	10.5	8.4
19	10.1	8.0	11.2	9.8	11.1	8.2
20	10.7	8.5	11.1	10.1	10.5	8.0
21	8.5	7.3	11.0	8.5	9.7	7.6
22	10.0	8.3	8.6	8.1	9.2	6.9
23	10.4	8.6	9.3	8.6	9.3	6.7
24	9.7	8.7	10.3	8.9	10.4	8.1
25	9.4	7.6	10.3	9.6	10.0	7.4
26	9.2	8.0	9.8	8.1	10.3	8.2
27	10.5	8.8	10.5	9.9	9.9	7.6
28	11.2	9.5	10.7	9.3	9.4	7.8
29	9.6	8.2	10.4	8.7	8.3	6.6
30	8.0	7.2	9.4	8.7	9.2	7.2

【实例辨析】 该资料的观测结果是糖化血红蛋白含量，为定量资料。研究者将 90 名患者随机分配到三组中，其试验设计类型是一个“假单因素三水平设计”，因为三个组实际上涉及两个因素，即“降糖药用否”和“盐酸二甲双胍片用否”，但缺少了一个“两种药都不用、仅有常规基础治疗组”。若简单地将三组视为“三种治疗方案(无法考察两药物之间是否存在协同和拮抗作用)”，可以将其近似视为“单因素三水平设计”。

另外，由于三组患者治疗前的糖化血红蛋白含量(协变量)影响各组药物治疗后的血红蛋白含量，且该定量因素的取值即使在同一组内部也不尽相同，因此为消除该协变量对试验效应的影响，可采用单因素三水平设计一元定量资料协方差分析。所需要的 SAS 程序如下：

```
data nkkysj2_5;
  do i=1 to 30;
    do c=1 to 3;
      input x y @@;
      output;
    end;
  end;
datalines;
10.8 9.4 10.4 9.2 9.8 7.6
11.6 9.7 9.7 9.1 11.2 7.9
10.6 8.7 9.9 8.9 10.7 9.0
9.0 7.2 9.8 8.6 9.6 7.8
```

```
11.2 10.0 11.1 9.9 10.1 8.5
9.9 8.5 8.2 7.1 9.8 7.5
10.6 5.3 8.8 7.8 10.1 8.3
10.4 8.1 10.0 7.9 10.3 8.2
9.6 8.5 9.0 8.0 11.0 8.4
10.5 9.1 9.4 9.0 10.5 8.1
10.6 9.2 8.9 7.9 9.2 7.0
9.9 8.4 10.3 8.9 10.1 7.7
9.5 7.6 9.3 8.9 10.4 8.0
9.7 7.9 9.2 8.1 10.0 6.6
10.7 8.8 10.9 10.2 8.4 6.1
9.2 7.4 9.2 8.5 10.1 8.1
10.5 8.6 9.2 9.0 9.3 7.8
11.0 9.2 10.4 8.9 10.5 8.4
10.1 8.0 11.2 9.8 11.1 8.2
10.7 8.5 11.1 10.1 10.5 8.0
8.5 7.3 11.0 8.5 9.7 7.6
10.0 8.3 8.6 8.1 9.2 6.9
10.4 8.6 9.3 8.6 9.3 6.7
9.7 8.7 10.3 8.9 10.4 8.1
9.4 7.6 10.3 9.6 10.0 7.4
9.2 8.0 9.8 8.1 10.3 8.2
10.5 8.8 10.5 9.9 9.9 7.6
11.2 9.5 10.7 9.3 9.4 7.8
9.6 8.2 10.4 8.7 8.3 6.6
8.0 7.2 9.4 8.7 9.2 7.2
;
run;
ods html;
proc glm;
  Class c;
  Model y=x c x* c/ssl;
run;
proc glm;
  Class c;
  Model y=x c /solution ss3;
  Lsmeans c / stderr tdiff pdiff;
run;
ods html close;
```

运行结果：

Source	DF	Type I SS	Mean Square	F Value	Pr > F
x	1	26.35220684	26.35220684	83.77	<.0001
c	2	19.74433114	9.87216557	31.38	<.0001
x*c	2	0.00161943	0.00080972	0.00	0.9974

以上是第一个过程步输出的结果，最后一行表明：三组之间由 X 推算 Y 的直线回归方程是平行的，即满足进行协方差分析的平行线假定。

Source	DF	Type III SS	Mean Square	F Value	Pr > F
x	1	27.91187132	27.91187132	90.83	<.0001
c	2	19.74433114	9.87216557	32.13	<.0001

以上是第二个过程步给出的第一部分结果，表明在试验设计时就关注了定量影响因素 X 对因变量 Y 的影响是正确的，消除其影响后，仍能发现三种处理之间校正的 Y 平均值之间的差别有统计学意义。

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	0.519001242	B	0.76128191	0.68	0.4972
x	0.722955639		0.07585700	9.53	<.0001
c 1	0.532119544	B	0.14352417	3.71	0.0004
c 2	1.147273561	B	0.14322919	8.01	<.0001
c 3	0.000000000	B	.	.	.

以上是第二个过程步输出的第二部分结果，呈现了线性模型中各参数的估计值、标准误的估计值、各参数与 0 之间差别是否具有统计学意义的假设检验结果。其中，模型的截距项与 0 之间的差别无统计学意义(严格地说，应将其删掉后重新分析，此处从略)。

c	y LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	8.25898851	0.10159490	<.0001	1
2	8.87414253	0.10145604	<.0001	2
3	7.72686896	0.10122418	<.0001	3

以上是第二个过程步输出的第三部分结果，呈现了三组 Y 校正平均值及其标准误的估计值以及各校正均值与 0 之间差别是否具有统计学意义的假设检验结果。

Least Squares Means for Effect c t for H0: LSMean(i)=LSMean(j) / Pr > t			
Dependent Variable: y			
i/j	1	2	3
1		-4.27147 <.0001	3.707526 0.0004
2	4.271474 <.0001		8.010054 <.0001
3	-3.70753 0.0004	-8.01005 <.0001	

以上是第二个过程步给出的第四部分结果，呈现了三组 Y 校正平均值之间两两比较的结果，结果表明：它们之间的差别均有统计学意义。

专业结论：患者治疗前的血红蛋白含量对试验效应的影响经协方差分析调整消除后，三个组治疗 3 个月的糖化血红蛋白含量不同。具体地说，两药联合使用，校正后的平均糖化血红蛋白含量最低；单用降糖药，校正后的平均糖化血红蛋白含量居中；而单用盐酸二甲双胍片，校正后的平均糖化血红蛋白含量最高。

【例 2-6】 某研究者在中老年人群尿白蛋白与肌酐比值的性别差异的研究中，比较了非糖尿病人群尿白蛋白与肌酐比值(ACR)和微量白蛋白尿(MAU)的测定结果，其中不同性别非糖尿病患者 Ucr 及 MAU 界值的比较结果见表 2-6。受试者的一般情况：去除糖尿病及肾脏病患者，共 1580 例完整资料纳入研究分析(平均年龄为 60 岁)，其中 NGT 有 1109 例，IGR 有 471 例，该人群与总人群年

龄及性别构成比差异无统计学意义。统计分析方法：各组计量资料用 $\bar{x} \pm s$ 表示，不符合正态分布的男女两组数据之间的比较采用非参数检验。ACR 正常值的判定以 95% 上位点具体数值为界点，大于界点判定为 MAU。统计结果：NGT、IGR 及合并后的非糖尿病人群中，男性 Ucr 明显高于女性；男性 ACR 均明显低于女性。试辨析资料所对应的设计类型。

表 2-6 不同性别非糖尿病人群 Ucr 及 MAU 界值的比较 ($M, P_5 - P_{95}$)

组 别	性 别	尿量 (mg/L)	Ucr (mmol/L)	MAU (mg/g)
非糖尿病	男	2.2(0 ~ 23.1)	8.1(2.4 ~ 20.7)	2.6(0 ~ 22.9)
	女	1.8(0 ~ 20.9)	5.5(1.8 ~ 16.9)	3.1(0 ~ 28.1)
NGT 组	男	2.2(0 ~ 21.5)	8.0(2.5 ~ 21.0)	2.5(0 ~ 22.5)
	女	1.9(0 ~ 18.8)	5.5(1.8 ~ 17.4)	3.2(0 ~ 27.0)
IGR 组	男	2.7(0 ~ 27.6)	8.1(2.4 ~ 18.9)	3.3(0 ~ 23.0)
	女	1.5(0 ~ 26.0)	5.5(1.8 ~ 14.6)	2.9(0 ~ 31.9)

【实例辨析】 该资料的观测指标是尿量 (mg/L)、Ucr (mmol/L) 和 MAU (mg/g)，均为定量观测指标。在本例资料中，研究者的分析方法有瑕疵，因为它按一个因素分层，对另一个因素各水平进行比较。事实上，资料中至少涉及两个因素，即“受试者种类”与“性别”，这两个因素都属于属性因素。没有一个标准试验设计类型与其对应，对数据进行分析时，可将其视为“具有两因素析因设计结构的一种模式”，并采用与其对应的统计分析方法。

2.2 配对设计与分析

2.2.1 应用目的

配对设计是一种特殊的设计方式，能够很好地控制非试验因素对结果的影响。配对设计的特点是，相同指标的观测结果是成对出现的，每对中的两个数据来自同一个或条件相近的两个个体。

2.2.2 设计方法

- 通常情况下，配对设计形式有三种：自身配对、同源配对、条件相近者配对。
- (1) 自身配对设计是对同一个体在处理前后或对同一个体的对称部位进行配对研究的方法。
 - (2) 同源配对设计是指亲代相同的两个个体(如同一窝别的条件相近的两只动物，性别相同的两个双胞胎等)分别服用不同的药物或接受不同的处理的方法。
 - (3) 条件相近者配对设计是指条件相近但并非同源的两个个体(如性别、体重、年龄、病情等相近者)接受不同的处理的设计方法。

2.2.3 具体设计过程

- (1) 先按主要非处理因素(如年龄、性别等)配对，然后为每个对子编号，如 1, 2, 3, ..., 并为每个对子中的两个受试对象编上编号，如 1 号和 2 号。
- (2) 为每个对子取随机数字。
- (3) 对每对中的受试对象随机分配处理，如处理 A 和处理 B。可事先规定，当某对子所对应的随机数为单数时，1 号接受 A 处理，2 号接受 B 处理；可事先规定，当某对子所对应的随机数为双数时，1 号接受 B 处理，2 号接受 A 处理；遇到 0 按双数看待。其随机分配结果见表 2-7。

表 2-7 配对设计随机分配结果表

研究对象对号	1	2	3	4	5	6	...	<i>n</i>
随机数	5	0	8	6	9	4	...	3
1 号归组	A	B	B	B	A	B	...	A
2 号归组	B	A	A	A	B	A	...	B

2.2.4 统计分析方法

对于配对设计定量资料，首先需要根据每对数据求其差值，如果差值服从或近似服从正态分布，就可以采用配对设计定量资料的 *t* 检验处理资料；如果差值不服从正态分布，可以考虑采用适当的方法使之满足正态性要求；然后采用配对设计定量资料的 *t* 检验处理资料，也可以考虑对差值直接用符号秩和检验进行统计分析。

对于配对设计的定性资料，根据研究目的选择合适的统计方法。当其研究目的是为了研究两种方法(如治疗方法)之间是否具有一致性时，应采用 Kappa 检验(即一致性检验)对其进行分析；当其研究目的是比较两种方法不一致部分的频数之间的差别是否具有统计意义时，可选用配对设计定性资料 χ^2 检验进行分析。

对于配对设计的等级资料，其本质是配对 2×2 列联表资料的“扩大”，即为配对 $R \times R$ 列联表资料。在处理此类资料时，人们主要关心的是两种方法(如治疗方法)之间的一致性，故采用 Kappa 检验对其进行分析。

2.2.5 SAS 实现程序及结果解释

【例 2-7】 对 12 份患者的血清分别用原方法(检测时间为 20min)和新方法(检测时间为 10min)检测谷 - 丙转氨酶，结果见表 2-8。问：两法所得结果有无差别？

表 2-8 12 份患者血清分别用原法和新法测谷 - 丙转氨酶 ($\text{nmol} \cdot \text{S}^{-1}/\text{L}$) 的比较

编 号	1	2	3	4	5	6	7	8	9	10	11	12
原法	60	142	195	80	242	220	190	25	198	38	236	95
新法	76	152	243	82	240	220	205	38	243	44	190	100

本资料属于配对设计四格表定量资料。先对其差值 *d* 进行正态性检验，若差值 *d* 服从正态分布，则采用配对 *t* 检验对其进行分析；若差值 *d* 不服从正态分布，则采用秩和检验进行分析。所需要的 SAS 程序如下：

```
data nkkysj2_7;
input x1 x2 @@;
d= x1 - x2;
datalines;
60 76
142 152
195 243
80 82
242 240
220 220
190 205
25 308
198 243
38 44
```

```
236      190
95      100
;
run;
ods html;
proc univariate normal;
var d;
run;
ods html close;
```

运行结果及其解释：

正态性检验			
检验	-----统计量-----		----- P 值-----
Shapiro-Wilk	W	0.875074	Pr < W 0.0758

正态性检验结果表明： $W = 0.875074$ ， $P = 0.0758 > 0.05$ ，说明差值 d 服从正态分布。故可采用配对设计一元定量资料 t 检验的结果。

位置检验： $\mu_0 = 0$

检验	-----统计量-----		----- P 值-----
学生 t	t	-1.3598	Pr > t 0.2011
符号	M	-3.5	Pr > = M 0.0654
符号秩	S	-21.5	Pr > = S 0.0566

配对设计一元定量资料 t 检验结果表明，其 $t = -1.3598$ ， $P = 0.2011 > 0.05$ ，故平均来说，原法和新法测谷 - 丙转氨酶的结果无差别。

需要注意的是：用配对设计定量资料差异性分析得出 $P > 0.05$ 结果后，千万不可认为两法测定结果是一致的！一致性检验应能反映每个个体上两法之间的吻合程度的高低。

【例 2-8】 对工厂 11 份工业污水测定氟离子浓度 (mg/L)，每份水样同时采用电极法及分光光度法测定，结果见表 2-9。问：就总体而言，这两种方法的测定结果有无差别？

表 2-9 两法测定某工厂的 11 份工业污水中氟离子浓度结果

样 品 号	氟离子浓度 (mg/L)	
	电 极 法	分光光度法
1	10.5	8.8
2	21.6	18.8
3	14.9	13.5
4	30.2	27.6
5	8.4	9.1
6	7.7	7.0
7	16.4	14.7
8	19.5	17.2
9	127.0	155.0
10	18.7	16.3
11	9.5	9.5

对表 2-9 中的定量资料采用自身配对设计一元定量资料差异性分析 (包括正态性检验) 所需要的 SAS 程序名为：nkksj2_8.sas。

```
data nkkysj2_8;
input x1 x2 @@;
d= x1-x2;
datalines;
  10.5      8.8
  21.6      18.8
  14.9      13.5
  30.2      27.6
   8.4      9.1
   7.7      7.0
  16.4      14.7
  19.5      17.2
 127.0     155.0
  18.7      16.3
   9.5      9.5
;
run;
ods html;
proc univariate normal;
var d;
run;
ods html close;
```

运行结果及其解释：

检验	正态性检验		----- P 值-----	
	-----统计量-----			
Shapiro-Wilk	W	0.456123	Pr < W	<0.0001

正态性检验结果表明： $W=0.456123$ ， $P < 0.0001$ ，说明差值 d 不服从正态分布，故采用秩和检验的结果。

位置检验： $\mu_0 = 0$

检验	-----统计量-----	----- P 值-----
学生 t	t -0.44084	Pr > t 0.6687
符号	M 3	Pr > = M 0.1094
符号秩	S 16	Pr > = S 0.1094

本例符号秩检验的结果：统计量 S 为 16，所对应的 $P=0.1094$ ，故就总体而言，这两种方法的测定结果无差别。

2.3 随机区组设计与分析

2.3.1 应用目的

随机区组设计(Randomized block design)，又称为配伍组设计。配伍因素常常是一种试图控制的非试验因素。配伍组设计是配对设计的扩大，当只有两个组别时即为配对设计。

2.3.2 设计方法

随机区组设计常有两种情况：第一种情况是同一受试对象在试验因素 K 个水平(可称为单个体

型区组)组间的比较(更确切地应称为具有一个重复测量的单因素设计);第二种情况是将受试对象按照某种特征(如动物的窝别、体重、年龄、性别等)相同或相近者组成 b 个单位组(配伍组或区组),每个单位组中有 k 个受试对象,分别随机分配到 k 个试验组中,然后比较各试验组间和区组间的效应,这种随机区组设计的特点是每个配伍组的例数等于试验因素的水平数,即试验组的个数。

2.3.3 具体设计过程

- (1) 试验单位编号,根据影响试验结果的非处理因素(如试验动物的性别、体重;病人的性别、年龄、病情等)配成 b 个区组,并依次编号。
- (2) 将每个区组中的 k 个受试对象分别随机地分配到 k 个处理组中。
- (3) 从随机数字表中读取随机数。
- (4) 对各区组内的随机数进行排序。
- (5) 根据区组内随机数序号决定处理类型。

【例 2-9】 将 12 个胃溃疡病人采用随机区组设计方案分配到 4 个处理组中,使其分别接受 A、B、C、D 四种药物的治疗。具体试验设计方案如下。

- ① 首先按照病人的溃疡面大小配成 3 区组,每个配伍组中各有 4 个溃疡面大小相似的病人,并依次编号: 1,2,3,4, 5,6, 7, 8, 9, 10,11,12。
- ② 从随机数字表中从任何一行、任何一列、任何一个数字开始、沿着任何方向依次取 12 个互不相同的 2 位随机数,如从第 10 行、第 1 列开始,沿着向右的方向依次读取随机数: 58,71,96,30, 24,18,46,23,34,27,85,13。
- ③ 对各区组内的随机数进行排序,排序后的顺序号见表 2-10。
- ④ 根据各区组内的随机数序号决定接受处理的药物种类: 序号为 1 的接受 A 药治疗,序号为 2 的接受 B 药治疗,序号为 3 的接受 C 药治疗,序号为 4 的接受 D 药治疗。

表 2-10 12 个胃溃疡病人的随机区组设计情况

区组号	1				2				3			
病人编号	1	2	3	4	5	6	7	8	9	10	11	12
随机数	58	71	96	30	24	18	46	23	34	27	85	13
序号(R)	2	3	4	1	3	1	4	2	3	2	4	1
处理类型	B	C	D	A	C	A	D	B	C	B	D	A

2.3.4 统计分析方法

随机区组设计与单因素 K 水平($K \geq 2$)设计相比,不仅考虑了试验因素的效应,还控制了一个非试验因素。当资料为服从正态分布且各总体方差相等的定量资料时,可采用与此设计对应的一元定量资料差异性分析(满足参数检验前提条件时选择相应设计一元定量资料方差分析,否则选择秩和检验)或者配对 t 检验(试验因素的水平数 = 2 时)进行不同组间的比较;当某些基线指标无法控制且对试验效应有影响时,需将该指标作为协变量,采用与此设计对应的一元定量资料协方差分析进行比较;当资料不满足正态性和(或)方差齐性时,可对原始数据进行转换或采用随机区组设计定量资料的秩和检验: 试验因素水平数 = 2 时,用 Wilcoxon 符号秩和检验;试验因素的水平数 ≥ 3 时,用 Friedman 秩和检验(M 检验)。

2.3.5 SAS 实现程序及结果解释

【例 2-10】 8 名受试对象在相同试验条件下分别接受 4 种不同频率声音的刺激,他们的反应率(%)资料见表 2-11。问: 4 种频率声音刺激的反应率是否有差别?

表 2-11 8 名受试对象对 4 种不同频率声音刺激的反应率(%) 比较

受 试 号	频 率 A	频 率 B	频 率 C	频 率 D
1	8.4	9.6	9.8	11.7
2	11.6	12.7	11.8	12.0
3	9.4	9.1	10.4	9.8
4	9.8	8.7	9.9	12.0
5	8.3	8.0	8.6	8.6
6	8.6	9.8	9.6	10.6
7	8.9	9.0	10.6	11.4
8	7.8	8.2	8.5	10.8

【实例辨析】 本例资料属于单个体型随机区组设计(严格地说,称为具有一个重复测量的单因素设计)一元定量资料,每个受试者均接受 4 种不同频率声音的刺激,即频率 A、频率 B、频率 C 和频率 D。在资料满足参数检验条件(假定各列数据之间互相独立)时,可近似采用随机区组设计一元定量资料的方差分析;若资料不满足参数检验条件,可采用 Friendman 秩和检验。下面列出随机区组设计一元定量资料方差分析(满足参数检验条件时)及 Friendman 秩和检验(不满足参数检验条件时)两种情况的 SAS 处理过程及其结果。

假设该资料满足参数检验的前提条件,则设实现随机区组设计一元定量资料方差分析所需要的 SAS 程序名为: nkkysj2_10. sas。

```
data nkkysj2_10;
  do a=1 to 8;
    do b=1 to 4;
      input x @@;
      output;
    end;
  end;
datalines;
  8.4    9.6    9.8    11.7
11.6    12.7   11.8   12.0
  9.4    9.1   10.4    9.8
  9.8    8.7    9.9   12.0
  8.3    8.0    8.6    8.6
  8.6    9.8    9.6   10.6
  8.9    9.0   10.6   11.4
  7.8    8.2    8.5   10.8
;
run;
ods html;
proc anova;
  class a b;
  model x = a b;
run;
ods html close;
```

程序说明: a 表示区组, b 表示 4 种频率。

运行结果:

Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	7	32.38000000	4.62571429	9.00	<.0001
b	3	14.38750000	4.79583333	9.34	<.0004

结果解释：

上述方差分析结果表明：分组变量 a 的 $F = 9.00$, $P < 0.0001$ ，说明各区组均数之间的差异有统计学意义。分组变量 b 的 $F = 9.34$, $P < 0.0001$ ，说明各试验组均数之间的差异有统计学意义。但各区组之间以及各实验组(各频率)之间的两两比较结果如何，还需进一步进行假设检验。下面仅给出各实验组之间采用 q 检验(即 SNK 检验)的结果：

Means with the same letter are not significantly different.			
SNK Grouping	Mean	N	b
A	10.8625	8	4
B	9.9000	8	3
B	9.3875	8	2
B	9.1000	8	1

SNK 检验所得结果表明，频率 A、频率 B 与频率 C 作用下的反应率之间的差别无统计学意义，这三种频率作用下的反应率低于频率 D 作用下的反应率。

假设该资料不满足参数检验条件，则 Friedman 秩和检验的 SAS 处理程序如下：

```
data nkkysj2_10;
input x a b @@;
datalines;
8.4 1 1 9.6 2 1 9.8 3 1 11.7 4 1
11.6 1 2 12.7 2 2 11.8 3 2 12.0 4 2
9.4 1 3 9.1 2 3 10.4 3 3 9.8 4 3
9.8 1 4 8.7 2 4 9.9 3 4 12.0 4 4
8.3 1 5 8.0 2 5 8.6 3 5 8.6 4 5
8.6 1 6 9.8 2 6 9.6 3 6 10.6 4 6
8.9 1 7 9.0 2 7 10.6 3 7 11.4 4 7
7.8 1 8 8.2 2 8 8.5 3 8 10.8 4 8
;
run;
Ods html;
proc freq;
tables b*a*x/scores =rank cmh2;
run;
ods html close;
```

运行结果：

Cochran-Mantel-Haenszel 统计量(基于秩得分)				
统计量	对立假设	自由度	值	概率
1	非零相关	1	15.0722	0.0001
2	行均值得分差值	3	15.1519	0.0017

总样本大小 = 32

结果说明：在输出窗口中输出的结果很多，关键结果为 CMH 检验结果。本例 CMH 统计量的第二个结果的值为 15.1519, $P = 0.0017 < 0.05$ ，说明各频率组反应率之间的差异有统计学意义。

2.3.6 实例辨析

【例 2-11】 某厂医务室测定 10 名氟作业工人工前、工中及工后 4h 的尿氟浓度($\mu\text{mol/L}$)。结果见表 2-12。问：氟作业工人在这三个不同时间的尿氟浓度有无差别？

表 2-12 10 名氟作业工人作业前后的尿氟浓度

工 人 编 号	工 前	工 中	工 后
1	90.53	142.12	87.38
2	88.43	163.17	65.27
3	47.37	63.16	68.43
4	175.80	166.33	210.54
5	100.01	144.75	194.75
6	73.69	138.96	200.02
7	46.32	126.33	65.27
8	105.27	126.33	100.01
9	86.32	121.06	105.27
10	60.01	73.69	58.95

【实例辨析】 对于本例资料，很多研究者往往误用单因素方差分析方法进行处理。其实，本例的研究者意在比较氟作业工人在三个不同时间的尿氟浓度的变化。此时，“时间”为处理因素，工前、工中及工后为其三种不同的水平，“工人”则是研究者为提高设计效率而采用的区组因素。因此，该资料符合随机区组设计(严格地说，称为具有一个重复测量的单因素设计)，可近似采用随机区组设计一元定量资料方差分析。

另外，本例资料中另一个容易使人混淆的处理因素是时间。重复测量是对每个受试对象的某项观察指标进行多次重复观测，而此处是对 10 名氟作业工人进行工前、工中、工后三次重复观测，故易让人联想到无对照的重复测量设计。这种看法并非没有道理，但作为一种科学的研究设计，我们必须保证其设计的严谨性。当研究者所关注的并非只是这三个时间的尿氟浓度的差异，而是随着时间的推移尿氟浓度的变动趋势时，可以考虑重复测量设计。

因此，在处理资料时，要结合资料的特点及其研究目的，进而选择恰当的统计学方法。下面给出该资料的 SAS 处理程序及其对结果的解释。

所需要的 SAR 程序如下：

```
data nkkysj2_11;
do a=1 to 10;
do b=1 to 3;
input x @@;
output;
end;
end;
datalines;
90.53 142.12 87.38
88.43 163.17 65.27
47.37 63.16 68.43
175.80 166.33 210.54
100.01 144.75 194.75
73.69 138.96 200.02
46.32 126.33 65.27
105.27 126.33 100.01
```

```
86.32    121.06    105.27
60.01     73.69     58.95
;
run;
ods html;
proc anova;
class a b;
model x = a b;
run;
ods html close;
```

程序说明：a 表示区组，b 表示三种处理方案。

运行结果：

The ANOVA Procedure					
Dependent Variable: x					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	47895.87729	4354.17066	4.51	0.0024
Error	18	17365.56086	964.75338		
Corrected Total	29	65261.43815			
		R - Square	Coeff Var	Root MSE	x Mean
		0.733908	28.27501	31.06048	109.8513
Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	9	39712.98388	4412.55376	4.57	0.0030
b	2	8182.89341	4091.44670	4.24	0.0310

结果解释：

上述方差分析结果表明： $F = 4.51$ ， $P = 0.0024$ ，说明模型有统计学意义。分组变量 a 的 $F = 4.57$ ， $P = 0.0030$ ，说明各区组均数之间的差异有统计学意义，即 10 名氟作业工人的尿氟浓度有差别；分组变量 b 的 $F = 4.24$ ， $P = 0.0310$ ，说明各处理组均数之间的差异有统计学意义，即工前、工中及工后 4h 的尿氟浓度之间也有差别。

2.4 析因设计及其一元定量资料的统计分析

2.4.1 应用目的

在医学实践中，需要了解各因素、各水平的作用，当一个因素的不同水平间随其他因素的水平不同而变化时，则称为因素间存在交互作用(interaction)。如果既想知道各因素的主效应，又要了解因素间的交互作用，这时需要应用一些较为复杂的试验设计，以便能够更准确地估计因素的主效应和各级交互作用效应的大小。能实现这一目的的试验设计方法有析因设计、正交设计、均匀设计等。其中，析因设计也称为全因子试验设计，是一种比较常见的设计方法，在试验研究中应用得比较多。当涉及的因素不超过 4 个或 5 个且每个因素的水平数也比较少时，可以考虑采用析因设计。

2.4.2 设计方法

析因设计是指将两个或多个因素的各水平进行全面组合，从而探讨各因素的作用及各因素间的交互作用。析因设计中所涉及的试验因素一般认为对观测结果的影响是地位平等的，并且在试验中同时施加。

1.2 个 2 水平试验因素析因设计(常简称为 2×2 析因设计)

在医学研究中,当试验因素有 2 个,其中每个因素有 2 个水平,2 个试验因素的水平全面组合、各组合条件下至少做 2 次独立重复试验,且由专业知识可知,2 个试验因素同时施加且对结果影响地位平等时,它们所决定的结构被称为 2×2 析因设计(或 2² 析因设计)。在 2×2 析因设计中共有 4(2×2=4)个试验组合,即人们习惯称呼的有 4 个试验组。

2. 多因素析因设计

在医学科研中,当试验因素有 2 个以上,且每一因素各有 2 个或 2 个以上水平,按一定排列方法将多个因素、多个水平全面组合起来,试验时全部试验因素同时施加且对结果的影响地位平等,这类试验安排被称为多因素析因设计。

2.4.3 具体设计过程

- (1)确定所研究的因素和水平,计算处理组合数。
- (2)确定试验设计模式:如有 A 与 B 两个因素,每个因素各有 2 个水平。则需采用 2×2 析因试验设计模式,表 2-13 是 2×2 析因设计模式。
- (3)处理组合的排列(快速有序,不重不遗)。

表 2-13 2×2 析因设计模式

A 因素	A、B 因素水平的组合情况	
	B 因素: 1 水平	2 水平
1 水平	A ₁ B ₁	A ₁ B ₂
2 水平	A ₂ B ₁	A ₂ B ₂

(4)当有交互作用存在时,必须对重复数做出估计,否则,对析因试验是没有意义的。重复次数估计有一个经验公式: $b = \frac{12}{a-1} + 1$, 其中 a 为组合数, b 为需要重复的次数。故当考察 A、B 两因素的交互作用时,试验需要重复的次数为 5 次。

(5)随机化:当处理组合数为 4,重复数为 5,共 4×5=20 次试验,共需 20 个受试对象。将受试对象随机分成 4 组。

2.4.4 统计分析方法

析因设计的统计分析,一般应检查定量资料是否满足参数检验的三个前提条件,即“独立性、正态性和方差齐性”。如满足条件,则可进行相应设计定量资料方差分析;若不满足条件,通常需要对资料做合适变量变换,仅当变换后的定量资料满足条件时再对其进行相应设计的定量资料方差分析。

2.4.5 SAS 实现程序及结果解释

1. 实例分析及 SAS 实现

【例 2-12】 将 20 例患有高胆固醇的患者随机分配到 4 个治疗组,第一组为基础治疗(既不用 A 药也不用 B 药),第二组在基础治疗的基础上用 B 药,第三组在基础治疗的基础上用 A 药,第四组在基础治疗上同时使用 A 药和 B 药,资料如表 2-14 所示。问:四种治疗方式的疗效之间的差别是否具有统计学意义? 两因素间交互作用是否具有统计学意义?

表 2-14 不同治疗组胆固醇的降低量 (mmol/L)

胆固醇的降低量		胆固醇的降低量	
不用 A 药(不用 B 药)	使用 B 药)	用 A 药(不用 B 药)	使用 B 药)
0.38	0.66	1.36	1.75
0.70	0.76	1.32	1.47
0.56	0.70	1.02	1.76
0.39	0.72	1.05	1.96
0.42	0.86	1.55	1.68

所需要的 SAS 程序如下：

```
data nkkysj2_12;
do i=1 to 5;
  do a=1 to 2;
    do b=1 to 2;
      input x @@;
      output;
    end;
  end;
end;
datalines;
0.38 0.66 1.36 1.75
0.70 0.76 1.32 1.47
0.56 0.70 1.02 1.76
0.39 0.72 1.05 1.96
0.42 0.86 1.55 1.68
;
run;
ods html;
proc anova;
class a b;
model x=a b a* b;
run;
ods html close;
```

运行结果：

Dependent Variable: x					
Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	1	3.84564500	3.84564500	145.56	<.0001
b	1	0.63724500	0.63724500	24.12	0.0002
a * b	1	0.05724500	0.05724500	2.17	0.1604

结果解释：

方差分析结果表明：

A 药用否与 B 药用否之间的交互作用无统计学意义($F=2.17$, $P=0.1604>0.05$),故有必要将交互作用项从模型中删除,获得精简的方差分析模型后再下结论。

删除交互作用项后的输出结果如下：

Source	DF	Anova SS	Mean Square	F Value	Pr > F
a	1	3.84564500	3.84564500	136.21	< .0001
b	1	0.63724500	0.63724500	22.57	0.0002
Level of	----- x -----				
a	N	Mean	Std Dev		
1	10	0.61500000	0.16847354		
2	10	1.49200000	0.30943676		
Level of	----- x -----				
1	10	0.87500000	0.44192131		
2	10	1.23200000	0.53416186		

本例 A 药物所对应的 $F = 136.21$, $P < 0.0001$, 表明 A 药物用与不用之间的差别有统计学意义, 由两水平下的平均值数值大小可知, 用 A 药比不用 A 药平均降低胆固醇值多, 其差量为 $1.492 - 0.615 = 0.877$; B 药物所对应的 $F = 22.57$, $P = 0.0002$, 表明 B 药物用与不用之间的差别有统计学意义, 由两水平下的平均值数值大小可知, 用 B 药比不用 B 药平均降低胆固醇值多, 其差量为 $1.232 - 0.875 = 0.357$ 。A 药与 B 药降低胆固醇量之差为 $1.492 - 1.232 = 0.26$ 。A 药作用稍强于 B 药, 但对于两者之间的差别是否有统计学意义, 未做假设检验, 不可武断地得出结论。

【例 2-13】 观察 A、B 两种镇痛药物联合运用在产妇分娩时的镇痛效果。A 药取 3 个剂量: 1.0mg, 2.5mg, 5.0mg; B 药也取 3 个剂量: 5 μ g, 15 μ g, 30 μ g。共 9 个处理组。将 27 名产妇随机等分为 9 组, 每组 3 名产妇, 记录每名产妇分娩时的镇痛时间, 结果见表 2-15。试分析 A、B 两种药物联合运用的镇痛时间。

表 2-15 A、B 两种药物联合运用的镇痛时间 (min)

A 药物剂量 (mg)	镇痛时间 (min)		
	B 药物剂量 (μ g) : 5	15	30
1.0	105	115	75
	80	105	95
	65	80	85
	75	125	135
2.5	115	130	120
	80	90	150
	85	65	180
5.0	120	120	190
	125	100	160

所需要的 SAS 程序如下：

```
data nkkysj2_13;
do a_dose=1.0,2.5,5.0;
  do rept=1 to 3;
    do b_dose=5,15,30;
      input x @@;
```

```
output;
end;
end;
end;
datalines;
105 115 75
80 105 95
65 80 85
75 125 135
115 130 120
80 90 150
85 65 180
120 120 190
125 100 160
;
run;
ods rtf;
proc glm;
class a_dose b_dose;
model x=a_dose|b_dose/ss3;
lsmeans a_dose* b_dose/tdiff pdiff;
run;
quit;
ods rtf close;
```

运行结果的第一部分：

Source	DF	Type III SS	Mean Square	F Value	Pr > F
a_dose	2	6572.222222	3286.111111	8.47	0.0026
b_dose	2	7022.222222	3511.111111	9.05	0.0019
a_dose*b_dose	4	7872.222222	1968.055556	5.07	0.0065

运行结果的第二部分：

a_dose	b_dose	x LSMEAN	LSMEAN Number
1	5	83.333333	1
1	15	100.000000	2
1	30	85.000000	3
2.5	5	90.000000	4
2.5	15	115.000000	5
2.5	30	135.000000	6
5	5	110.000000	7
5	15	95.000000	8
5	30	176.666667	9

这是两因素各水平组合下定量观测指标的平均值和对应的编号。

运行结果的第三部分：

Least Squares Means for Effect a_dose*b_dose									
t for H0: LSMean(i)=LSMean(j) / Pr > t									
Dependent Variable: x									
i\j	1	2	3	4	5	6	7	8	9
1		-1.03633 0.3138	-0.10363 0.9186	-0.41453 0.6834	-1.96903 0.0645	-3.21263 0.0048	-1.65813 0.1146	-0.72543 0.4775	-5.80346 <.0001
2	1.036333 0.3138		0.9327 0.3633	0.6218 0.5419	-0.9327 0.3633	-2.1763 0.0431	-0.6218 0.5419	0.3109 0.7594	-4.76713 0.0002
3	0.103633 0.9186	-0.9327 0.3633		-0.3109 0.7594	-1.8654 0.0785	-3.109 0.0061	-1.5545 0.1375	-0.6218 0.5419	-5.69983 <.0001
4	0.414533 0.6834	-0.6218 0.5419	0.3109 0.7594		-1.5545 0.1375	-2.7981 0.0119	-1.2436 0.2296	-0.3109 0.7594	-5.38893 <.0001
5	1.969032 0.0645	0.9327 0.3633	1.865399 0.0785	1.554499 0.1375		-1.2436 0.2296	0.3109 0.7594	1.243599 0.2296	-3.83443 0.0012
6	3.212632 0.0048	2.176299 0.0431	3.108998 0.0061	2.798099 0.0119	1.243599 0.2296		1.554499 0.1375	2.487199 0.0229	-2.59083 0.0185
7	1.658132 0.1146	0.6218 0.5419	1.554499 0.1375	1.243599 0.2296	-0.3109 0.7594	-1.5545 0.1375		0.9327 0.3633	-4.14533 0.0006
8	0.725433 0.4775	-0.3109 0.7594	0.6218 0.5419	0.3109 0.7594	-1.2436 0.2296	-2.4872 0.0229	-0.9327 0.3633		-5.07803 <.0001
9	5.803464 <.0001	4.767131 0.0002	5.69983 <.0001	5.388931 <.0001	3.834431 0.0012	2.590832 0.0185	4.145331 0.0006	5.078031 <.0001	

结果解释：

本例 A 因素的 $F = 8.47$, $P = 0.0026$, 说明 A 药不同剂量镇痛作用的差别有统计学意义；B 因素的 $F = 9.05$, $P = 0.0019$, 说明 B 药不同剂量镇痛作用的差别也有统计学意义；A 和 B 因素的交互项 A * B 的 $F = 5.07$, $P = 0.0065$, 说明两种药物剂量在镇痛效果上存在交互作用，在上面的“编号”中，编号为 6 和编号为 9 的两种剂量搭配下镇痛效果为比较好和最好的两种剂量搭配，它们之间的差别仍有统计学意义，说明当两种药物均取最高浓度时，镇痛效果最佳。必须强调的是，应注意安全性！

2. 案例辨析与释疑

【例 2-14】 某研究者欲观察老年患者心肌缺血 (MI) 和左心室肥大 (LVH) 时心脏自主神经张力的变化情况，对 73 例老年冠心病 (CHD)、高血压性心脏病 (HHD) 和原发性高血压 (EH) 患者与 30 例健康老年人的心率变异性进行了分析。方法：使用 NHK-2000 型综合心电图检测仪 (南京)，对研究对象的 12 导联心电图采样进行时域分析。分析观察指标主要有 R-R 间期的标准差 (SDNN)、相邻 R-R 间期差值的均方根 (rMSSD)、相邻 R-R 间期差值 $\geq 50\text{mm}$ (PNN50) 所占的百分比和心率变异指数 (HRVI) 等，结果见表 2-16。统计学处理：参数以 $\bar{x} \pm s$ 表示，各组间比较用 t 检验。试辨析该实例存在的错误。

表 2-16 患者组与对照组 HRV 测定结果及比较 ($\bar{x} \pm s$)

组 别	例 数	SDNN (ms)	rMSSD (ms)	PNN 50 (%)
患者组	73	$26 \pm 11^{\Delta}$	$17 \pm 10^{\Delta}$	$4 \pm 6^{\Delta}$
CHD 组	38	26 ± 13	17 ± 9	4 ± 5
HHD 并 EH 组	18	25 ± 13	17 ± 10	4 ± 4

续表				
组 别	例 数	SDNN (ms)	rMSSD (ms)	PNN 50 (%)
CHD 合并 HHD 和 EH 组	17	27 ± 11	16 ± 7	4 ± 5
对照组	30	34 ± 13	31 ± 23	14 ± 20

注:与对照组比较,*代表 $P < 0.05$,[△]代表 $P < 0.01$,其余 $P > 0.05$;患者组各分组间比较均为 $P > 0.05$ 。

【实例辨析】 研究者将其视为单因素设计定量资料,采用 t 检验对资料进行处理。这种做法分隔了资料的整体性,无法全面反映资料的特点,且某些组间的差别是由多个因素造成的。此资料的合适做法是依据专业知识及统计知识对资料进行合理拆分,拆分结果如下。

组合 1:患者组、对照组。

组合 2:CHD 组、HHD 并 EH 组、CHD 合并 HHD 和 EH 组。

组合 1 只涉及 1 个两水平的试验因素,为成组设计定量资料。在资料满足参数检验的前提下,可选用成组设计定量资料 t 检验;若组间不满足方差齐性,但满足正态性时可采用 t' 检验或秩和检验;若资料不满足正态性时,可采用 Wilcoxon 秩和检验。

组合 2 涉及 1 个三水平的试验因素(疾病种类),在资料满足参数检验的前提下,可选用单因素三水平设计一元定量资料方差分析进行处理;若组间不满足参数检验条件,可采用秩和检验进行分析。

【例 2-15】 某研究者欲探讨伊贝沙坦和福辛普利对老年原发性高血压患者血管内皮功能的保护作用。将 62 例 1 个月内未服用过血管紧张素转换酶抑制剂和血管紧张素受体拮抗剂类药物的老年原发性轻、中度高血压患者随机分为福辛普利组和伊贝沙坦组,治疗 3 个月。用药前、后分别行肱动脉血管内皮依赖性舒张功能(FDD)超声检测,血浆一氧化氮(NO)、内皮素(ET-1)和血管性假血友病因子(vWF)水平测定。观察并比较两组、治疗前后 FDD、血浆 NO、ET-1 和 vWF 水平的变化。结果见表 2-17。统计学处理:试验结果均采用均数 ± 标准差表示,应用 SPSS10.0 统计软件包处理数据,两组用药前或用药后组间比较采用独立样本的 t 检验;每组用药前后的组内比较,采用配对 t 检验。试辨析存在的错误。

表 2-17 两组患者用药前后血浆中 NO、ET-1、vWF 的变化($\bar{x} \pm s$)

组 别	例 数	NO(μmol/L)	ET-1 (nmol/L)	vWF (%)
福辛普利组用药前	31	15.09 ± 3.91	28.04 ± 7.83	85.29 ± 9.71
用药 12 周后	31	19.10 ± 5.84 [#]	21.76 ± 7.56 [#]	75.32 ± 7.91 [#]
伊贝沙坦组用药前	31	15.22 ± 3.72	27.72 ± 8.23	86.29 ± 9.82
用药 12 周后	31	20.44 ± 5.75 [#]	22.01 ± 6.53 [#]	74.94 ± 7.85 [#]

注:与治疗前比较,*代表 $P < 0.05$,[#]代表 $P < 0.01$ 。

【实例辨析】 本例资料含有两个试验因素:一个试验因素是组别(药物种类),含有两个水平,分别为福辛普利组和伊贝沙坦组;另一个试验因素是时间,有两个水平,分别为用药前和用药 12 周后。故该资料属于具有一个重复测量两因素设计的定量资料,建议采用 SAS 中的 MIXED 过程并基于具有一个重复测量两因素设计的定量资料统计分析。

2.5 正交设计

2.5.1 应用目的

正交试验设计法最早由日本质量管理专家田口玄一提出,称为国际标准型正交试验法。应用正交设计试验方案可以根据生物学和医学的专业知识,只分析有意义因素的主效应和部分重要因素的交互作用,从而有效减少试验因素的水平组合数。

2.5.2 设计方法

正交设计是按照正交表和相应的交互表进行设计，对多因素、多水平进行非全面试验，不仅可了解各因素的作用程度，而且能知道各因素之间存在的低阶交互作用，能找出诸因素各水平的最佳搭配。

2.5.3 具体设计过程

1. 正交设计具体过程

正交设计是非全面试验，是析因设计的部分实施。其设计步骤如下：

- (1) 根据研究目的，确定观察指标(指标的取值应满足方差分析的条件)。
- (2) 拟定观察指标所有可能的影响因素及其水平数，选择合适的正交表。

2. 正交表

正交表是正交设计的主要工具，具有安排多因素及筛选主要因素的功能，根据正交表可以安排各因素各水平的部分组合。每个正交表都有一个表头符号，记为 $L_N(m^k)$ ，表示该正交表有 N 行 k 列，每一列由整数 $1, 2, \dots, m$ 组成。用表 $L_N(m^k)$ 安排试验时， N 表示试验次数， k 表示最多可以安排的因素个数， m 表示各因素的水平数。例如表 2-18 为 $L_8(2^7)$ 正交设计表。该表有 8 个行，表示可安排 8 种处理；7 个列，表示可以安排 7 个不同的因素，而 2 则表示每列所安排的因素水平数最多为 2。

表 2-18 $L_8(2^7)$ 正交设计表

处 理 编 号	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

正交表的主要特征可以总结为：每一列中不同数字出现的次数相同，即整齐可比性；任两列不同数对出现的次数相同，即均匀分散性。这两条决定了正交表所确定的试验点具有数学上的“正交性”。

选择正交表的原则如下：

- (1) 各试验因素的水平数最好相等。
- (2) 操作简单的试验或希望得到较多的信息，可选择 N 较大的正交表。反之，操作复杂或成本较高的试验，可选择 N 较小的正交表。
- (3) 根据专业知识确定各因素间是否存在交互作用及要观察的交互项：交互作用一般只考虑一级交互作用，当因素间交互作用多时，宜选择列多的正交表；交互作用项少时，选择列较少的正交表。
- (4) 确定试验次数：受研究因素、因素水平及时间、经费等因素的影响，试验次数不宜太多或太少。对研究因素和研究水平已经确定的试验，可通过选择具有相同因素水平，不同试验次数的正

交表来增加试验，也可以通过进行重复试验来增加试验次数，以便更准确地估计试验误差，提高试验精度。

值得一提的是，通常，正交表上有两类误差：其一，来自正交表的空列(它反映的是不同试验条件之间的试验误差，统计学上常称为准确性)；其二，来自各相同试验条件下的重复试验(它反映的是试验结果的重现性好坏，统计学常称为精确性)。一个正交设计下的试验研究，最好同时具有前述两类试验误差，共同构成整个正交设计的试验误差。基于这样的试验误差的自由度进行的统计分析结果更稳定。

(5)综合以上情况进行表头设计。当表头设计中涉及交互作用时，应根据相应的交互效应表交互作用项安排于相应的列中。

(6)按照表头设计的要求安排试验，收集数据。

2.5.4 统计分析方法

正交设计的统计分析与析因设计的分析方法相似，从略。

2.5.5 SAS 实现程序及结果解释

1. 无重复试验的正交设计

【例 2-16】 某研究者欲研究雌螺产卵的最优条件，在 20cm² 的泥盒里饲养同龄雌螺 8 只，试验条件有 4 个因素，即温度、含氧量、含水量、pH 值，每个因素 2 个水平，见表 2-19。试验的具体安排和试验结果见表 2-20。若考虑温度与含氧量对雌螺产卵有交互作用，试对该试验结果进行统计分析。

表 2-19 雌螺产卵条件因素与水平

试 验 序 号	A 因素温度(℃)	B 因素含氧量(%)	C 因素含水量(%)	D 因素 pH 值
1	5	0.5	10	6.0
2	25	5.0	30	8.0

表 2-20 雌螺产卵条件的正交试验(安排在 1、2、4、7 列上)

试 验 序 号	A 因素温度(℃)	B 因素含氧量(%)	C 因素含水量(%)	D 因素 pH 值	产 卵 数 量
1	5	0.5	10	6.0	86
2	5	0.5	30	8.0	95
3	5	5.0	10	8.0	91
4	5	5.0	30	6.0	94
5	25	0.5	10	8.0	91
6	25	0.5	30	6.0	96
7	25	5.0	10	6.0	83
8	25	5.0	30	8.0	88

值得一提的是，正交表各行试验条件下不做独立重复试验的前提条件是：正交表中至少有一个空列；若各行上做多次独立重复试验，其误差大小在专业上允许的范围之内。否则，正交设计的结果是不稳定的，即试验结果不具有重现性。

所需要的 SAS 程序如下：

```
data nkkysj2_16;
  input A B C D x @@;
  datalines;
```

```
5 0.5 10 6.0 86
5 0.5 30 8.0 95
5 5.0 10 8.0 91
5 5.0 30 6.0 94
25 0.5 10 8.0 91
25 0.5 30 6.0 96
25 5.0 10 6.0 83
25 5.0 30 8.0 88
;
run;
ods html;
proc anova;
    class A B C D;
    model x=A B C D A* B ;
run;
ods html close;
```

运行结果：

Source	DF	Anova SS	Mean Square	F Value	Pr > F
A	1	8.00000000	8.00000000	3.20	0.2155
B	1	18.00000000	18.00000000	7.20	0.1153
C	1	60.50000000	60.50000000	24.20	0.0389
D	1	4.50000000	4.50000000	1.80	0.3118
A * B	1	50.00000000	50.00000000	20.00	0.0465

结果说明：从各因素的方差分析结果可知，只有含水量对雌螺产卵数量有影响，其他因素均无影响。温度和含氧量之间有交互作用。

此时，应从 *P* 值最大的项开始删除，逐渐精简方差分析模型。

Source	DF	Anova SS	Mean Square	F Value	Pr > F
A	1	8.00000000	8.00000000	2.53	0.2102
B	1	18.00000000	18.00000000	5.68	0.0973
C	1	60.50000000	60.50000000	19.11	0.0222
A*B	1	50.00000000	50.00000000	15.79	0.0285

说明温度与含氧量之间的交互作用、含水量对产卵数量的影响有统计学意义。可以尝试将 A 因素淘汰掉后再运行程序，此处从略。

Level of C	N	x	
		Mean	Std Dev
10	4	87.7500000	3.94757309
30	4	93.2500000	3.59397644

Level of A	Level of B	N	x	
			Mean	Std Dev
5	0.5	2	90.5000000	6.36396103
5	5	2	92.5000000	2.12132034
25	0.5	2	93.5000000	3.53553391
25	5	2	85.5000000	3.53553391

由上述的平均值可知，含水量为 30% 的产卵数量多于含水量为 10% 的产卵数量；含氧量分别为 0.5% 与 5% 且当温度为 5℃ 时，产卵数量相近（含氧量高时，产卵数量稍高一点）；而当温度为 25℃ 时，含氧量低时的产卵数量反倒要多一些。

2. 有重复试验的正交设计

【例 2-17】 过氧乙酸是一种广泛应用的杀灭肝炎病毒的消毒剂，但其有效成分极不稳定，以致影响消毒效果。可通过试验找出有关因素（表 2-21）对其稳定性的影响，指出主要因素及各因素所起的作用，并选出各因素共同作用的最佳搭配，成为保护过氧乙酸稳定性的最优条件。要求研究 A、B、C、D 四因素的独立作用，并根据经验和专业知识确定的交互作用 A × B，A × D。试验的具体安排和试验结果见表 2-22。

表 2-21 过氧乙酸稳定性试验的因素及水平

试 验 因 素	表 示 符 号	水 平 1	水 平 2
稳定剂(0.3% 磷酸)	A	加	不加
水浴温度(℃)	B	25 ~ 30	35 ~ 40
浸泡口表	C	10 支	不浸
加盖	D	加	不加

表 2-22 过氧乙酸稳定性试验安排及其结果

试 验 号	不同因素(列号)的水平						放置 24h 过氧乙酸残存量	
	A(1)	B(2)	C(4)	D(7)	A × B(3)	A × D(6)	x ₁	x ₂
1	1	1	1	1	1	1	7.00	4.11
2	1	1	2	2	1	2	6.05	3.50
3	1	2	1	2	2	2	1.10	0.80
4	1	2	2	1	2	1	1.90	0.96
5	2	1	1	2	2	1	2.40	1.65
6	2	1	2	1	2	2	4.00	1.50
7	2	2	1	1	1	2	0.35	0.30
8	2	2	2	2	1	1	0.30	0.90

注：第 5 列为空列，各行决定的试验条件下做了两次独立重复试验。

所需要的 SAS 程序如下：

```
data nkkysj2_17;
  input A B C D @@;
  do i=1 to 2;
    input x @@;
    output;
```

```
end;
datalines;
1 1 1 1 7.00 4.11
1 1 2 2 6.05 3.50
1 2 1 2 1.10 0.80
1 2 2 1 1.90 0.96
2 1 1 2 2.40 1.65
2 1 2 1 4.00 1.50
2 2 1 1 0.35 0.30
2 2 2 2 0.30 0.90
;
run;
ods html;
proc anova;
  class A B C D;
  model x=A B C D A*B A*D;
run;
ods html close;
```

运行结果：

Source	DF	Anova SS	Mean Square	F Value	Pr > F
A	1	12.28502500	12.28502500	9.27	0.0139
B	1	34.81000000	34.81000000	26.27	0.0006
C	1	0.12250000	0.12250000	0.09	0.7680
D	1	0.73102500	0.73102500	0.55	0.4765
A * B	1	4.20250000	4.20250000	3.17	0.1086
A * D	1	0.16402500	0.16402500	0.12	0.7330

此结果表明，还有 4 项无统计学意义，有必要将 P 值最大的两项先从方差分析模型中删除，当发现模型中还有较大 P 值的项时继续删除，最终得如下结果：

Source	DF	Anova SS	Mean Square	F Value	Pr > F
A	1	12.28502500	12.28502500	9.32	0.0093
B	1	34.81000000	34.81000000	26.40	0.0002

方差分析检验结果表明，A 因素（稳定剂， $F = 9.32$ ， $P = 0.0093$ ）和 B 因素（水浴温度， $F = 26.40$ ， $P = 0.0002$ ）是影响过氧乙酸稳定性的两个主要因素。

Level of A	N	x	
		Mean	Std Dev
1	8	3.17750000	2.40101615
2	8	1.42500000	1.28730061

Level of B	N	x	
		Mean	Std Dev
1	8	3.77625000	1.97833218
2	8	0.82625000	0.5388595

由平均值可知，加稳定剂放置 24h 过氧乙酸残留量高；水浴温度低时放置 24h 过氧乙酸残留量高。

2.5.6 实例辨析与释疑

【例 2-18】 原文题目：环氧合酶抑制剂 NS-398 抑制食管癌干细胞增强放射敏感性的研究。方法：采用四甲基偶氮唑盐(MTT)法，将照射后的细胞立即制成单细胞悬液常规接种，同时设空白组(无细胞培养液)。分别于 24h、48h 及 72h 终止培养，在酶联免疫检测仪上选择 490nm 滤光片测定吸光度(A)值，并记录结果，计算出细胞生长抑制率。结果见表 2-23。统计学处理：采用 SPSS 1.5 统计软件进行统计分析，对酶标仪检测 A_{490} 值的统计分析， A_{490} 值值用均数 \pm 标准差($\bar{x} \pm s$)表示，采用正交设计资料方差分析。试辨析存在的错误。

表 2-23 NS-398 联合不同剂量的 X 线对 Eca-109 细胞的生长抑制作用 ($\bar{x} \pm s, n=6$)

照射剂量	24h			48h			72h		
	A_{490} 值		IR (%)	A_{490} 值		IR (%)	A_{490} 值		IR (%)
	试 验 组	对 照 组		试 验 组	对 照 组		试 验 组	对 照 组	
0	0.48 \pm 0.02	0.33 \pm 0.04	31.7	0.68 \pm 0.02	0.38 \pm 0.04	44.7	0.87 \pm 0.04	0.48 \pm 0.02	45.1
2	0.38 \pm 0.04	0.22 \pm 0.03	42.5	0.29 \pm 0.03	0.16 \pm 0.02	45.4	0.20 \pm 0.02	0.09 \pm 0.04	53.3
4	0.27 \pm 0.04	0.12 \pm 0.03	54.1	0.19 \pm 0.03	0.09 \pm 0.01	53.9	0.13 \pm 0.02	0.05 \pm 0.03	64.2
6	0.20 \pm 0.03	0.08 \pm 0.01	57.5	0.10 \pm 0.02	0.04 \pm 0.00	60.9	0.07 \pm 0.02	0.03 \pm 0.00	64.8
8	0.15 \pm 0.01	0.05 \pm 0.01	63.8	0.08 \pm 0.03	0.03 \pm 0.00	65.9	0.03 \pm 0.00	0.01 \pm 0.00	63.4

注： A_{490} 为波长为 490 nm 的吸光度值；IR 为抑制率。

【实例辨析】 本例资料中的观测指标为波长为 490nm 的吸光度值。一个因素是照射剂量，它是一个重要的试验因素，有 5 个水平；表头上还涉及两个试验因素：一个试验因素是时间有 3 个水平，即 24h、48h 和 72h；另一个试验因素是“组别”，即试验组和对照组，包含 2 个水平。原作者声称采用“正交设计资料方差分析”，这种说法不恰当。因为一个涉及 5、3、2 水平的三个试验因素的研究场合下，目前尚无与之对应的正交设计表。该资料的正确处理方法为：若三个试验因素对结果的影响有主次之分，则为三因素嵌套设计；若三个试验因素对结果的影响无主次之分，则为三因素析因设计。

【例 2-19】 某研究者采用正交设计的方法分析醇沉影响因素，用高效液相法测定蒙花苷的含量，研究清热去湿片稠膏醇沉过程中，在确定乙醇加入量和静置时间的前提下，考察其他因素对蒙花苷转移率的影响，其正交试验因素水平表见表 2-24，试验结果见表 2-25。试辨析存在的错误。

表 2-24 正交试验因素水平表

水 平	A: 加入速度 (ml/min)	B: 搅拌时间 (min)	C: 静置温度 (℃)
1	20	10	5
2	50	20	25
3	100	40	40

表 2-25 正交试验 $L_9(3^4)$ 结果

试 验 序 号	各因素下的水平代码			蒙花苷转移率 (%)
	加入速度 (A)	搅拌时间 (B)	静置温度 (℃)	
1	1	1	1	63.9
2	1	2	2	68.2

续表				
试 验 序 号	各因素下的水平代码			蒙花苷转移率(%)
	加入速度(A)	搅拌时间(B)	静置温度(℃)	
3	1	3	3	67.3
4	2	1	2	69.8
5	2	2	3	68.9
6	2	3	1	60.8
7	3	1	3	70.5
8	3	2	1	64.9
9	3	3	2	64.3

【实例辨析】 本例资料中的观测指标为蒙花苷转移率。含有3个试验因素，各有3个水平，因此该试验应为“三因素三水平试验”。针对该资料，若研究者想全面了解加入速度、搅拌时间和静置温度3个因素各水平对蒙花苷转移率的作用，宜选用析因设计。原作者为了减少试验组合数(最本质的目的是希望减少总试验次数)，选择了一个行数最少的三水平正交表来安排此试验，各行上未做重复试验。虽然正交表中还有一个空列，似乎可以采用正交设计定量资料的方差分析处理资料，但结果的可信度较低。因为不能绝对地排除3个试验因素之间不存在任何交互作用，也不能肯定各行试验条件下的试验误差一定在专业上允许的范围之内。所以在以上两个方面没有确切把握时，建议采用析因设计安排试验为宜。

2.6 拉丁方设计

2.6.1 应用目的

拉丁方设计方法可控制两个非试验因素，进而增加均衡性，减少试验误差，提高试验效率。另外，拉丁方设计也可大大减少试验次数，因而在一些经费昂贵的动物试验和试验室研究中且满足一定的条件下，可考虑选用拉丁方设计类型来安排试验。

2.6.2 设计方法

用 r 个拉丁方字母排成 r 行 r 列的方阵，使每行每列中每个字母都只出现一次，这样的方法称为 r 阶拉丁方或 $r \times r$ 拉丁方。按拉丁方的字母安排试验因素，按行和列安排重要非试验因素的试验称为拉丁方试验。需要注意的是，拉丁方设计虽然比随机区组设计多安排了一个已知的对试验结果有影响的非试验因素，但这两种试验都是单因素(指试验因素)试验。

拉丁方试验设计的要求有：必须是3个因素的试验，且3个因素的水平数相等；3个因素间均无交互作用或交互作用在本质上可忽略不计。另外，最关键的还有受试对象被观测的次数，若每个受试对象需要被重复观测 r 次，则被称为单个体型拉丁方设计，最节省样本含量，但其结果很可能有严重问题，此时，结果的准确程度取决于试验因素对评价指标的影响程度。例如，若试验因素是秤的种类或体温计的种类，它们对被称物品的重量或被测生物体的体温几乎没有什么影响，此时，采用单个体型拉丁方设计，可称为“花小钱办大事”，结果准确可信；若试验因素是血压计的种类，它对被测生物体的血压没有实质性的影响但会受到心理因素的暂时影响，此时，用单个体型拉丁方设计，其结果的准确度就会有所降低；若试验因素是药物的种类或药物剂量，在生物体活体试验中，它们对被检测的生物体的主要评价指标有实质性的影响，此时，若仍采用单个体型拉丁方设计，其结果的准确度是非常低的，甚至会得出错误的结果和结论，此种情形下应慎用或不用单个体型拉丁方设计。

2.6.3 具体设计过程

拉丁方设计是在随机区组设计的基础上发展起来的一种设计方法，其设计方法与随机区组设计方法有所不同。

1. 具体设计步骤

- (1) 根据试验目的、试验因素及其水平数等选择恰当规模的拉丁方。
- (2) 将试验因素的 r 个水平随机地排列成 $r \times r$ 拉丁方阵。
- (3) 决定采用单个体型还是多个体型拉丁方设计。
- (4) 若是单个体型的，则将 r 个受试对象随机地安排在 r 行上；若是多个体型的，则将 $r \times r$ 个个体按某重要非试验因素形成 r 个区组，每个区组内有 r 个条件相近的个体，随机决定每个区组所在的行。
- (5) 在纵向位置(即列)上安排第二个重要非试验因素的各个水平。该因素若是时间，一般不进行随机化处理，而是按自然的先后顺序排列。

2. 实例分析

某研究者为比较甲、乙、丙、丁、戊 5 种体温计测定人体的体温，选用 5 例高血压患者进行试验，对每例患者采用 5 种体温计重复测定 5 次，其设计步骤如下：

- (1) 本例有 5 种不同的体温计，5 例高血压患者，每例患者被重复测定 5 次，故可选 5×5 基本拉丁方。
- (2) 对基本方阵进行随机排列。例如，先将第 2 行和第 3 行对调，然后将第 1 行和第 4 行对调，见图 2-1 第 1 行。
- (3) 随机排列拉丁方的列。例如，在第 (2) 步的基础上先将第 2 列和第 4 列对调，然后将第 1 列和第 3 列对调，见图 2-1 第 2 行。
- (4) 随机安排试验因素的水平：事先假定拉丁方阵中的字母 A、B、C、D、E 对应丁、戊、甲、丙、乙 5 种体温计；即得字母 $A \rightarrow$ 丁、 $B \rightarrow$ 戊、 $C \rightarrow$ 甲、 $D \rightarrow$ 丙、 $E \rightarrow$ 乙。

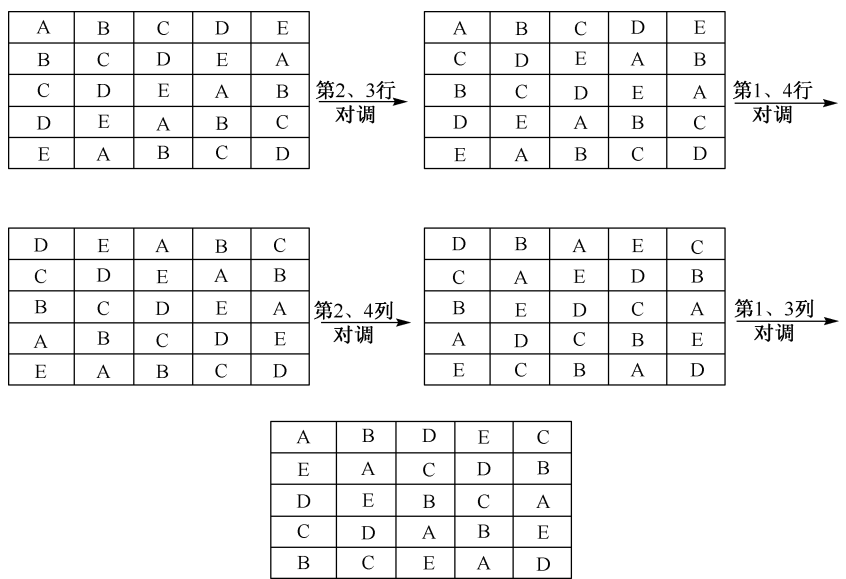


图 2-1 基本拉丁方行列随机变换流程图

(5)将上述最后一个拉丁方的行、列和拉丁字母分别对应于患者、测定顺序和体温计种类，并按该拉丁方安排试验，见表 2-26。

表 2-26 拉丁方设计随机分配结果

患 者 编 号	体温计代号				
	* : I	II	III	IV	V
1	丁	戊	乙	丙	甲
2	乙	丁	甲	乙	戊
3	丙	丙	戊	甲	丁
4	甲	乙	丁	戊	丙
5	戊	甲	丙	丁	乙

注：* 代表测定顺序。

2.6.4 统计分析方法

1. 实例分析及结果解释

【例 2-20】 某研究者为了比较甲、乙、丙、丁、戊、己 6 种药物给人体注射后产生的皮肤疱疹大小(mm²)，采用拉丁方设计，将 36 名志愿者按年龄、性别等因素配伍成 6 个区组，每个区组 6 个研究对象(此乃多个体型拉丁方设计)，并在每个研究对象的 1 个特定部位进行注射。试验结果见表 2-27。试做统计分析。

表 2-27 6 种药物的多个体型拉丁方设计及试验结果

志愿者区组 (行区组因素)	药物代号(皮肤疱疹大小, mm ²)					
	* : 1	2	3	4	5	6
1	C(87)	B(75)	E(81)	D(75)	A(84)	F(66)
2	B(73)	A(81)	D(87)	C(85)	F(64)	E(79)
3	F(73)	E(73)	B(74)	A(78)	D(73)	C(77)
4	A(77)	F(68)	C(69)	B(74)	E(76)	D(73)
5	D(64)	C(64)	F(72)	E(76)	B(70)	A(81)
6	E(75)	D(77)	A(82)	F(61)	C(82)	B(61)

注：* 代表注射顺序(列区组因素)。

所需要的 SAS 程序如下：

```
data nkkysj2_20;
  do r=1 to 6;
    do c=1 to 6;
      input z $ x @@;
      output;
    end;
  end;
datalines;
C 87  B 75  E 81  D 75  A 84  F 66
B 73  A 81  D 87  C 85  F 64  E 79
F 73  E 73  B 74  A 78  D 73  C 77
A 77  F 68  C 69  B 74  E 76  D 73
D 64  C 64  F 72  E 76  B 70  A 81
E 75  D 77  A 82  F 61  C 82  B 61
;
```

```
run;
ods html;
proc anova;
class r c z;
model x=r c z;
run;
ods html close;
```

运行结果：

Source	DF	Anova SS	Mean Square	F Value	Pr > F
r	5	250.4722222	50.0944444	1.47	0.2447
c	5	85.4722222	17.0944444	0.50	0.7723
z	5	667.1388889	133.4277778	3.91	0.0124

结果解释：

方差分析结果表明，在本例中，因素 r 对应的 $F=1.47$ ， $P=0.2447$ ，表明 6 个配伍组研究对象产生皮肤疱疹大小的总体均数之间的差异无统计学意义；因素 c 对应的 $F=0.50$ ， $P=0.7723$ ，表明 6 个注射顺序对应的皮肤疱疹大小的总体均数之间的差异也无统计学意义；因素 z 对应的 $F=3.91$ ， $P=0.0124$ ，表明 6 种药物产生皮肤疱疹大小的总体均数之间的差异有统计学意义。此时，应将无统计学意义的两个因素删除，重新运行程序。修改后 SAS 程序的过程步如下：

```
ods html;
proc anova;
class z; model x=z; means z/snk;
run;
ods html close;
```

修改后输出的结果如下：

Source	DF	Anova SS	Mean Square	F Value	Pr > F
z	5	667.1388889	133.4277778	3.93	0.0074

此时，对药物因素(z)各水平对结果影响的评价更合理、更稳定(误差项的自由度由原先的 20 增加为现在的 30)， $F=3.93$ ， $P=0.0074$ ，说明不同药物所产生的平均皮肤疱疹面积大小不同。

Means with the same letter are not significantly different.				
SNK Grouping		Mean	N	z
	A	80.500	6	A
	A	77.333	6	C
	A	76.667	6	E
B	A	74.833	6	D
B	A	71.167	6	B
B		67.333	6	F

由 q 检验(或称为 SNK 检验)可知，A 药与 F 药副作用之间的差别有统计学意义，A 药引起的

皮肤疱疹面积大，而 F 药引起的皮肤疱疹面积小。其他任何两种药物副作用之间的差别无统计学意义。

2.6.5 实例辨析与释疑

【例 2-21】 某研究者在胰岛素效价测定研究中，采用拉丁方试验设计方案，对 4 例糖尿病患者随机分配 4 种处理(A 试剂低剂量、A 试剂高剂量、B 试剂低剂量、检验品 B 试剂高剂量)，使每例患者接受处理的机会相等，并在 4 个不同的时间点测量血糖，试验结果见表 2-28。试辨析其存在的错误。

表 2-28 4 例糖尿病患者胰岛素效价测定试验结果

患 者 编 号	血糖值反应量(10mg/L)			
	第 1 次	第 2 次	第 3 次	第 4 次
1	102.9	84.7	130.3	114.3
2	77.8	91.5	107.5	144.0
3	84.7	81.2	128.0	112.0
4	68.7	97.2	107.5	139.4

【实例辨析】 从表 2-28 的纵向 4 列可知，涉及两个试验因素，即处理类别(A 试剂品与 B 试剂品)与剂量高低(低与高)，同一行上的四个数据来自同一例患者，故属于具有两个重复测量两因素设计，而不是拉丁方设计。

值得一提的是，后两列数据明显大于前两列数据，若四列数据对应的因素水平组合分别依次为 A 试剂低剂量、A 试剂高剂量、B 试剂低剂量、B 试剂高剂量，则表明：A 试剂与 B 试剂之间的差别明显大于剂量之间的差别，当此结果有专业依据且具有重现性时，可以认定此资料所对应的设计类型为含区组因素的嵌套设计。

2.7 交叉设计与分析

2.7.1 设计方法

交叉设计(cross-over design)是在自身配对设计的基础上发展起来的三因素设计方法，指受试对象在不同的试验阶段分别接受试验因素两个不同水平的作用。交叉设计的试验效应主要受到以下几个因素的影响：试验因素、试验阶段、试验顺序等因素。最简单的交叉设计是试验因素的水平数为 2，试验顺序因素也为 2 的 2×2 配对交叉设计或成组交叉设计。

2.7.2 具体设计过程

将研究对象随机分为甲、乙两组，甲组在第 I 阶段给予 A₁ 处理，在第 II 阶段给予 A₂ 处理；乙组在第 I 阶段给予 A₂ 处理，在第 II 阶段给予 A₁ 处理，这是成组交叉设计。

例如，为检查 A 和 B 两台代谢测定器测定耗氧量的结果是否相同，采用交叉试验设计方案，对条件相似的 14 名健康人进行测试。具体试验设计方案如下。

- (1)将 14 名健康人按完全随机方案随机均分为 2 组，即甲组和乙组。
- (2)在第 I 阶段，对甲组的 7 名患者先用 A 仪器测定，对乙组的 7 名患者先用 B 仪器测定；在第 II 阶段，对甲组的 7 名患者再用 B 仪器测定，对乙组的 7 名患者再用 A 仪器测定。
- 按照交叉试验方案，对每个研究对象都用 A、B 两台仪器各测定 1 次。

2.7.3 SAS 实现程序及结果解释

【例 2-22】 在评价两种剂量的阿司匹林对患者胃出血的 2×2 交叉试验中，试验数据见表 2-29。试分析两种剂量的阿司匹林效果之间的差异是否具有统计学意义。

表 2-29 阿司匹林对患者胃出血的效应

患者编号	次序	时期 1	时期 2
1	AB	5.1	3.8
2	AB	2.9	3.9
3	BA	2.7	3.2
4	BA	3.5	3.2
5	AB	4.4	5.8
6	BA	4.1	4.0
7	AB	6.2	6.4
8	BA	1.6	2.3
9	AB	5.3	4.7
10	BA	3.9	4.2
11	BA	3.7	4.5
12	AB	3.2	3.2
13	AB	2.4	2.8
14	BA	4.3	3.7
15	AB	5.2	3.4
16	BA	2.6	2.4
17	BA	4.2	3.0
18	AB	2.1	2.6
19	BA	4.4	5.2
20	AB	1.7	2.9

所需要的 SAS 程序如下：

```
data nkkysj2_22;
  do r=1 to 20;
    do time=1 to 2;
      input treat $ x @@;
      output;
    end;
  end;
datalines;
A 5.1 B 3.8
A 2.9 B 3.9
B 2.7 A 3.2
B 3.5 A 3.2
A 4.4 B 5.8
B 4.1 A 4.0
A 6.2 B 6.4
B 1.6 A 2.3
A 5.3 B 4.7
B 3.9 A 4.2
```

```
B 3.7 A 4.5
A 3.2 B 3.2
A 2.4 B 2.8
B 4.3 A 3.7
A 5.2 B 3.4
B 2.6 A 2.4
B 4.2 A 3.0
A 2.1 B 2.6
B 4.4 A 5.2
A 1.7 B 2.9
;
run;
ods html;
proc anova;
class r time treat;
model x = r time treat;
run;
ods html close;
```

运行结果及其解释：

Source	DF	Anova SS	Mean Square	F Value	Pr > F
r	19	47.12275000	2.48014474	6.36	0.0001
time	1	0.07225000	0.07225000	0.19	0.6720
treat	1	0.00225000	0.00225000	0.01	0.9403

方差分析结果表明：仅个体的结果之间差别有统计学意义，其他因素对结果的影响均无统计学意义。

2.8 重复测量设计

2.8.1 设计方法

所谓重复测量是指同一个受试对象在接受某种处理后，在某个因素不同水平条件下或某几个因素水平组合条件下被重复观测两次或两次以上的方法。通常情况下所指的重复测量资料是指同一受试对象在不同的时间点上多次测量同一个指标所得到的资料，但有时与重复测量有关的试验因素也可以是受试对象的两个或多个对称的部位。另外，在有些试验研究中，如果是将来自某个个体的某种组织或细胞分成了几份，然后再对这几份进行几种不同的处理，此时“处理”则为与重复测量有关的试验因素。重复测量资料最突出的一个特点是，测量同一受试对象所得的多个数据之间往往具有不等的相关性，即距离越近的数据之间的相关性越强，反之亦然。

2.8.2 具体设计过程

一般地说，很多试验都可以采用重复测量设计来安排，即在试验过程中定期记录观察结果。常见的重复测量类型有前后测量设计、对称部位上的重复测量设计等。

在 K 个不同时间点上对每个受试者的同一个定量指标进行重复观测，资料表达形式如表 2-30 所示。

表 2-30 具有一个重复测量两因素设计一元定量资料

受试者编号	干预因素 A	观测指标 (单位)	
		时间因素 T: T_1	T_k
1	A_1	X	X
2	A_1	X	X
\vdots	A_1	X	X
m	A_1	X	X
m + 1	A_2	X	X
m + 2	A_2	X	X
\vdots	A_2	X	X
2m	A_2	X	X

注:“X”代表指标的具体测定值。

2.8.3 统计分析方法

重复测量设计的定量资料统计方法很多,若资料满足正态性、方差齐性和球对称性,则可采用相应设计类型的一元方差分析;若资料不满足球对称性的条件,则可选用校正的一元方差分析、多元方差分析及混合效应模型分析等。由于重复测量数据具有自相关特点,所以对该类资料进行统计分析时,需注意对数据的协方差结构予以考虑。

2.8.4 SAS 实现程序及结果解释

【例 2-23】 某研究者在某项研究中,从甲厂与乙厂各随机收集了 10 名职工连续三年(2007—2009 年)的甘油三脂 (mmol/L)测定结果,数据见表 2-31。

表 2-31 甲厂与乙厂各 10 名职工 2007—2009 年甘油三脂测定值

分 组	编 号	甘油三脂 (mmol/L)		
		2007 年	2008 年	2009 年
甲厂	1	1.820	1.160	0.790
	2	0.481	0.730	0.900
	3	0.572	0.700	1.610
	4	0.503	1.030	1.120
	5	0.653	0.880	0.850
	6	1.100	1.110	1.430
	7	0.765	0.500	2.950
	8	3.330	3.770	3.810
	9	1.670	3.250	2.790
	10	0.543	2.200	1.280
乙厂	1	8.000	5.450	7.000
	2	2.770	2.500	1.120
	3	1.090	1.760	1.200
	4	3.910	6.170	5.280
	5	0.950	1.670	0.700
	6	0.980	0.790	0.500
	7	0.990	1.490	1.710
	8	0.570	1.890	0.800
	9	1.200	1.340	1.100
	10	1.420	1.760	1.200

本研究对象来自两个厂(甲厂和乙厂),每个受试对象重复测量3次,故按照设计类型,该资料符合具有一个重复测量的两因素设计资料,它是医学研究应用中最常见且较简单的一种。但由于计算机分析应用技术普及程度较低,计算烦琐,研究者常分别对各时点值进行单因素方差分析或进行简单的多元方差分析,而忽视了重复测量数据误差的多层次性和数据间存在的自相关性,给实际问题的解释带来了不必要的麻烦。

该类重复测量资料提倡采用 SAS 中的 MIXED 过程进行处理分析,具体的 SAS 程序如下:

```
data nkkysj2_23;
input group id time y;
datalines;
1 1 1 1.820
1 2 1 0.481
1 3 1 0.572
1 4 1 0.503
1 5 1 0.653
1 6 1 1.100
1 7 1 0.765
1 8 1 3.330
1 9 1 1.670
1 10 1 0.543
2 1 1 8.000
2 2 1 2.770
2 3 1 1.090
2 4 1 3.910
2 5 1 0.950
2 6 1 0.980
2 7 1 0.990
2 8 1 0.570
2 9 1 1.200
2 10 1 1.420
1 1 2 1.160
1 2 2 0.730
1 3 2 0.700
1 4 2 1.030
1 5 2 0.880
1 6 2 1.110
1 7 2 0.500
1 8 2 3.770
1 9 2 3.250
1 10 2 2.200
2 1 2 5.450
2 2 2 2.500
2 3 2 1.760
2 4 2 6.170
2 5 2 1.670
2 6 2 0.790
2 7 2 1.490
2 8 2 1.890
2 9 2 1.340
2 10 2 1.760
```

```
1 1 3 0.790
1 2 3 0.900
1 3 3 1.610
1 4 3 1.120
1 5 3 0.850
1 6 3 1.430
1 7 3 2.950
1 8 3 3.810
1 9 3 2.790
1 10 3 1.280
2 1 3 7.000
2 2 3 1.120
2 3 3 1.200
2 4 3 5.280
2 5 3 0.700
2 6 3 0.500
2 7 3 1.710
2 8 3 0.800
2 9 3 1.100
2 10 3 1.200
;
ods html;
proc mixed data=nkkysj2_23 method
=ml maxiter=50;
class group id;
model y = group time time (group)/
noint solution;
repeated/ type=un sub=id r rcorr;
run;
ods html close;
```

程序说明: id 为研究对象编号, group 为工厂分组(分为甲厂 =1, 乙厂 =2); time 为测定时间(2007 年 =1,2008 年 =2,2009 年 =3); y 代表测定变量(甘油三脂)。

运行结果及其解释:

Fit Statistics		
-2 Log Likelihood		144.0
AIC (smaller is better)		194.0
AICC (smaller is better)		233.4
BIC (smaller is better)		201.5

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
20	77.02	<.0001

以上两部分输出结果为模型拟合的统计量结果值。

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
group	2	10	24.82	0.0001
time	1	10	0.05	0.8244
time (group)	1	10	28.87	0.0003

以上结果表明，甲厂和乙厂两组人群(各 10 人)甘油三脂测定结果的平均水平有差别($F = 24.82$, $P = 0.0001$)；不同测定时间的甘油三脂测定结果之间无差别($F = 0.05$, $P = 0.8244$)；测定时间和厂别之间存在交互作用($F = 28.87$, $P = 0.0003$)。

2.8.5 实例辨析与释疑

【例 2-24】 为评价万古霉素对残肾功能的影响，某研究者将 30 名腹膜炎患者随机分成两组：万古霉素组(腹腔使用去甲万古霉素，每次 1 袋 2000mL 透析液，15 ~ 30 mg/kg，5 ~ 7 天 1 次)和头孢唑啉组(每次 1 袋透析液中加入 1g，每天 1 次)。其中间计算结果见表 2-32。试对比分析两组患者感染前后的尿量和残肾功能差值(ARRF)的变化。试辨析设计类型。

表 2-32 两组在腹膜炎感染半年前后的尿量和 RRF 差值($\bar{x} \pm s$)

项 目	万古霉素组		头孢唑啉组	
	感 染 前	感 染 后	感 染 前	感 染 后
尿量	835.8 ± 234.7	697.4 ± 219.0	928.8 ± 294.9	765.4 ± 247.5
RRF 差值	2.89 ± 1.88	2.23 ± 1.49	2.45 ± 2.01	2.19 ± 1.36

【实例辨析】 首先从该资料的表达形式看，该资料的表达不够规范，应将试验分组因素放在表的左边，观测指标放在表头短横线上边。另外，该资料从设计及分析角度而言，研究者将 30 名腹膜炎患者随机等分成两组(一组使用万古霉素，另一组使用头孢唑啉)，两组患者在试验条件上的差别在于所用的药物种类不同，故药物种类是一个试验因素。此外，研究者在感染前后两个时间点监测患者的尿量和 RRF 值，故检测时间是一个重复测量因素。所以本例资料类型为具有一个重复测量两因素设计定量资料。

【例 2-25】 某研究者欲探讨血管紧张素Ⅱ受体拮抗剂(ARB)依贝沙坦能否延缓腹膜透析患者残存肾功能(RRF)的丢失。方法：将入选的 48 例病情稳定的维持性腹膜透析患者随机分为依贝沙坦组和对照组，依贝沙坦组予以安博维 300mg/d。统计学方法：所有数据统计采用 SPSS 11.0 统计软件包进行分析。计量资料采用 $\bar{x} \pm s$ 表示，组间比较采用方差分析。统计结果见表 2-33。结论：研究结束时，依贝沙坦组和对对照组的收缩压和舒张压、血红蛋白、血钾水平差异均无统计学意义；依贝沙坦组肌酐清除率(CCL)高于对照组；两组 24h 尿量均减少，但对照组较依贝沙坦组减少明显。试辨析该资料的设计类型。

表 2-33 依贝沙坦组和对对照组患者的临床资料($\bar{x} \pm s$)

组 别	时 间	尿量 (mL/d)	收缩压 (mmHg)	舒张压 (mmHg)	CCL[L.周 ⁻¹ . (1.73m ²) ⁻¹]	Hb(g/L)	K ⁺ (mmol/L)
依贝沙坦	研究前	1304 ± 395	133.0 ± 13.2	83.0 ± 12.3	66.2 ± 14.0	88 ± 13	3.95 ± 0.43
	研究后	885 ± 276	134.0 ± 15.7	81.0 ± 11.6	63.0 ± 16.9	89 ± 15	4.19 ± 0.52

续表							
组 别	时 间	尿量 (mL/d)	收缩压 (mmHg)	舒张压 (mmHg)	CCL[L·周 ⁻¹ · (1.73m ²) ⁻¹]	Hb(g/L)	K ⁺ (mmol/L)
对照	研究前	1206 ± 459	135.0 ± 15.8	83.0 ± 10.8	67.8 ± 14.6	87 ± 15	3.97 ± 0.34
	研究后	663 ± 312	137.0 ± 16.4	80.0 ± 12.5	59.0 ± 14.8	88 ± 16	4.06 ± 0.31

【实例辨析】 研究者将 48 例病情稳定的维持性腹膜透析患者随机分为依贝沙坦组和对照组，依贝沙坦组予以安博维 300 mg/d，对照组不进行处理，两组患者在试验条件上的差别在于是否给予依贝沙坦，故药物是一个试验因素。此外，研究者在研究前后两个时间点监测患者的尿量、收缩压、舒张压、血红蛋白、血钾等临床资料，故检测时间是一个重复测量因素。所以本例资料类型为具有一个重复测量的两因素设计定量资料。

(石福艳 王素珍 孟维静)

第3章 内科试验设计与数据分析

3.1 内科试验设计的三要素

3.1.1 内科试验设计概述

内科试验设计是在专业理论的指导下,围绕人类身心健康,对尚未研究或深入研究的事物进行探讨,旨在揭示事物矛盾的内部联系与客观规律,较准确地回答和解决所提出的新观点、新技术。随着对内科学科研的重视,科研设计的地位越来越高,作用越来越大。

依照研究对象的不同,内科学的试验研究通常分为动物试验、临床试验和社区干预试验。

动物试验是指以试验动物为试验对象(或称受试对象,或称试验材料),采用研究者关心的试验因素,观察试验因素不同水平对试验动物的效应或作用大小。

动物试验主要分为以下三类。

(1)确证性试验:用一种最大可能,正确解答问题的系统方法,收集有关处理实施后动物反应的科学数据。

(2)探索性试验:为产生新的假说提供素材。

这两类试验都涉及在试验人员的控制下,对动物施加一定的处理或操作,来发现是否试验因素的不同在受试对象身上引起不同的反应;如有可能,需对这种反应进行定量检测。

(3)检验性试验:是观察性的试验,用于发现科研人员不能控制的变量之间的联系。这种联系可能是暂时的,也可能是长期的。

临床试验(Clinical Trial),指任何在人体(病人或健康志愿者)进行药物的系统性研究,以证实或揭示试验药物的作用、不良反应及/或试验药物的吸收、分布、代谢和排泄,目的是确定试验药物的疗效与安全性。临床试验一般分为I、II、III和IV期临床试验。

社区试验(community trial),也称为社区干预项目(community intervention program, CIP)、生活方式干预试验(lifestyle intervention trial)、以社区为基础的公共卫生试验(community-based public health trial)等。是以尚未患所研究疾病的人群作为整体进行试验观察,常用于对某种预防措施或方法进行考核或评价。社区试验接受干预的基本单位是整个社区,有时也可以是某一人群的各个亚群。如果某种疾病的危险因子分布广泛,不易确定高危人群时,也需要采用社区试验。

3.1.2 试验设计的三要素

内科学试验研究的基本要素包括影响因素(包括试验因素和非试验因素)、受试对象和试验效应。如用某种传统西药或中成药治疗缺铁性贫血,观察比较两组试验动物血红蛋白的上升趋势,该研究中所用的两种药物称为试验因素,患有缺铁性贫血的试验动物称为受试对象,血红蛋白取值的高低称为试验效应。

1. 影响因素

影响因素通常指由外界施加于受试对象的因素,包括生物的、化学的、物理的或内外环境的。但是,生物本身的某些特征(如遗传特性、心理因素等)也可作为试验因素(研究者特别关心的)来

进行观察，通常将它们视为重要非试验因素(研究者不想关心，但它们对试验结果确有影响)更贴切。因此，研究者应正确、恰当地确定影响因素。一般应注意以下几点。

(1)抓住试验研究中的主要因素。研究中的主要因素是按照在以往研究基础上(本人或他人)提出的某些假设和要求来决定的。一次试验涉及的试验因素不宜太多，否则会使分组增多，受试对象的例数增多，在实施中难以控制误差。然而，试验因素过少，又难以提高试验的广度和深度。因此，需根据研究目的的需要与实施的可能来确定带有关键性的试验因素。

【例 3-1】 要改进某种细胞的培养方法，与其有关的因素很多，如温度、pH 值、培养液、培养时间等。其中，每个因素又分若干水平(或等级)，如温度从 34℃ 至 38℃，每 1℃ 为一个水平，则有 5 个水平；pH 值从 6.5 至 7.4，每 0.1 为一个水平，则有 10 个水平。当培养液有两个水平；培养时间有 3 个水平时，应如何合理地选择处理因素和各个因素的水平数？

【分析解答】 根据题目中给出的结果，总共有四个因素，各个因素水平数各不相同，如表 3-1 所示。如果每个因素的每个水平都要考虑到的话，则须做 $5 \times 10 \times 2 \times 3 = 300$ 种条件的试验；若每种条件下试验重复 10 次的话，就需要做 3000 次试验，不可能在一次或几次试验中完成。

【实例辨析】 本试验的主要目的是研究两种不同的培养液对细胞培养的作用，因此可以将培养液作为主要的研究因素，而其他的如温度、pH 值、培养时间等，则不是主要的研究因素；结合专业知识，该细胞的培养温度在 37℃ 时效果最好，因此将温度恒定在 37℃；而 pH 值对该细胞的影响较小，因此采用中性溶液，即 pH 值为 7 即可，因此，在众多因素与水平中抓住主要的因素，且因素的水平数不宜过多。

(2)找出重要非试验因素。除了确定的试验因素以外，凡是影响试验结果的其他因素都称为非试验因素，所产生的混杂效应也影响了试验因素产生的效应对比和分析。例如，利用两种不同药物治疗某种细菌感染的试验，非试验因素可能有试验动物的种类、属性、喂养方式等。如果两组试验动物的种类、属性、喂养方式等构成不一，则可能影响药物疗效的比较。因此，设计时便设法控制这些非试验因素，只有这样才能消除它们的干扰作用，减少试验误差。

表 3-1 四个因素及其水平	
因 素	水 平 数
温度	5(34 ~ 38℃)
pH 值	10(6.5 ~ 7.4)
培养液	2
培养时间	3

(3)试验因素必须标准化。试验因素的强度、频率、持续时间与施加方法等，都要通过查阅文献和预试验找出各自的最适条件，然后制定出有关规定和制度，并使之相对固定，否则会影响试验结果的评价。如试验因素是药物种类，必须正确选择批号，给药途径和时间也应标准化和相对固定化。

2. 受试对象

受试对象的选择十分重要，对试验结果有着极为重要的影响。研究课题不同，对动物的要求也往往不同。动物的选择应该注意种类、品系、年龄、性别、体重、窝别和营养状况等，有时还要考虑动物的生活环境等。

试验动物的选择是医学研究中首先应予考虑的问题。选择恰当与否关系到课题的质量、经费开支、研究途径及方法的简繁，甚至影响到课题的成败及研究结果的正确性。医学科研中，作为受试对象的前提是，所选对象必须同时满足两个基本条件：必须对试验因素敏感；反应必须稳定。对动物的选择应该充分考虑动物本身对试验的影响，主要包括以下几个方面。

(1)首先是种属。动物的种属非常重要，应根据种属之间的差异正确选择试验动物，否则可能贻误整个试验。

实例 1：鼠、小鼠、豚鼠和兔对催吐药不产生呕吐反应，而猫、犬和人则容易产生呕吐。

实例2：衍生物对犬、猫、豚鼠等均能引起与人相似的病理变化，产生变性血红蛋白，但对家兔则不易产生变性血红蛋白，对小鼠则完全不产生。

(2)品系的影响。应根据动物品系的差异正确选择试验动物。

实例1：BA/2 小鼠发生听源性癫痫为 100%；而 C57BL 小鼠根本不出现这种反应。

实例2：B/cAnN 小鼠对放射线极敏感，而 C57BR/CdN 小鼠对放射线却具有抗力。

实例3：L/N 小鼠对疟原虫易感，而 C58/LwN 小鼠对疟原虫感染有抗力。

实例4：“诱发性排卵动物”，一般情况下只有交配才引起排卵，与人及其他一些哺乳动物的自发性排卵有较大差异，在应用试验结果时应注意。

(3)应该考虑动物的年龄和体重。一般来说，幼年动物比成年动物敏感，一些慢性试验，因观察时间较长，可选择年幼、体重较小的动物做试验。如研究性激素对机体影响的试验，一定要用幼年或新生的动物。老年动物的代谢功能低下，反应不灵敏，一般不选用。进行一些老年医学的研究应选用老年动物。咖啡碱对老年大鼠的毒性较大，对幼年大鼠毒性较小。一般动物试验设计应选成年动物。

【例3-2】 为了测定戊烷、苯和二氯乙烷等急性毒性试验的 LD₅₀及麻醉浓度并观察其敏感性，幼年>老年>成年，如表3-2 所示。

表 3-2 成年动物的年龄、体重和寿命比较

	小 鼠	大 鼠	豚 鼠	兔	犬
成年日龄(d)	65 ~ 90	85 ~ 110	90 ~ 120	120 ~ 180	250 ~ 360
成年体重(g)	20 ~ 28	200 ~ 280	350 ~ 600	2000 ~ 3500	8000 ~ 15000
平均寿命(y)	1 ~ 2	3 ~ 3	5 ~ 6	13 ~ 17	
最高寿命(y)	>3	>4	>6	>13	>34

(4)应该考虑动物的性别。科研工作中，一般优先选用雄性动物或雌雄各半做试验，动物性别对试验结果不影响的试验或一定要选用雌性动物的试验例外。

【例3-3】 3 月龄的 Wistar 大鼠摄取乙醇量按单位体重计算，雌性比雄性多，雌的排泄量也多，表3-3 表示的是药物反应性的性别差异。

表 3-3 药物反应性的性别差异

药 物	动 物 种	感受性强的性别	药 物	动 物 种	感受性强的性别
肾上腺素	大鼠	雄	巴比妥酸盐类	大鼠	雌
乙醇	小鼠	雄	哇巴因	大鼠	雄
四氧嘧啶	小鼠	雌	印防己毒素	大鼠	雌
氨基比林	小鼠	雄	钾	大鼠	雄
新肿凡纳明	小鼠	雌	硒	大鼠	雌
铅	大鼠	雄	海葱	大鼠	雌
野百合碱	大鼠	雄	固醇类激素	大鼠	雌
烟碱	小鼠	雄	士的宁	大鼠	雌
氨基蝶呤	小鼠	雄	碘胺	大鼠	雌
乙苯基	大鼠	雌	地辛	犬	雄
苯	家兔	雌	二硝基苯酚	猫	雌
四氯化碳	大鼠	雄	麦角固醇	小鼠	雄
氯仿	小鼠	雄	乙基硫氨酸	大鼠	雌

(5)应该考虑的是动物的生理状态。动物的生理状态(如怀孕、哺乳时),对外界环境因素作用的反应性常较不怀孕、不哺乳的动物有较大差异。因此,在一般试验研究中不宜采用这种动物。

动物的功能状态不同,也影响对药物的反应。如动物在体温升高的情况下对解热药比较敏感,而体温正常时对解热药就不敏感;血压高时对降压药比较敏感,而在血压低时对降压药敏感性就差,反而可能对升压药比较敏感。

(6)应该考虑的是动物的健康状态。一般情况下,健康动物对药物的耐受量比有病的动物要大,所以有病动物比较易于中毒死亡。患有疾病或处于衰竭、饥饿、寒冷、炎热等条件下的动物会影响试验结果,因此一定要选用健康动物进行试验。

实例1:动物发炎组织对肾上腺激素的血管收缩作用极不敏感。

实例2:营养条件差的动物不易复制成功动脉粥样硬化动物模型。

实例3:动物发热可使代谢增加,体温升高 1°C ,代谢率增加7%。

动物潜在感染,对试验结果影响很大,造成试验中断或试验失败。

实例4:寄生虫、细菌和病毒感染等会对试验结果有影响。不应有小鼠脱脚病(鼠痘),病毒性肝炎和肺炎、伤寒;鼠的沙门菌病、病毒性肺炎、化脓性中耳炎;豚鼠VC缺乏症、传染性肺炎、沙门菌病;兔的球虫病、巴氏杆菌病;犬的狂犬病、犬瘟热;猫的传染性白细胞减少症、肺炎;猕猴的结核病、肺炎、痢疾等。

3. 试验效应

一般是通过某些观察指标,定量或定性地反映试验效应。研究者应当对欲研究的问题有较为全面的了解,在试验设计中千万不要遗漏了某些重要的观察指标,以免试验结果不可靠。

(1)选择观察指标的基本原则。

①客观性:客观性指标是指通过精密设备或仪器测定的数据,能真实显示试验效应的大小或性质;排除了人为因素的干扰。在选择观察指标时应尽可能选择客观指标,避免一些笼统的、不确切的指标。有时,客观指标还具有判断的主观性问题,如X线胸片是客观的,但判断时存在主观性问题,所以,对于这种情况,须制定明确的判断标准。

②精确性:选用的指标应尽量精确。指标的精确性包括准确性和精确性,准确性是指所观察结果的真实程度,即观测值与真值的接近程度,属系统误差;精确性是指所观察结果的深度,即重复观测时,观测值与平均值的接近程度,属随机误差。试验效应指标既要准确又要精密,而首先是准确可靠。

③灵敏性:应尽量选择高灵敏性的指标,即选择能够显著提高灵敏性的仪器对观察指标进行测量。但也要考虑试验经费问题,灵敏度高的仪器的费用往往昂贵,应根据试验经费,选择相对廉价且灵敏度高的测量方法。

【例3-4】采用动物试验研究某药治疗缺铁性贫血的效果,如何选择指标?

【实例辨析】采用动物模型做试验,既可选用动物的症状、体征,也可选用血红蛋白含量等作为观察指标,但这些指标均不够灵敏,只有在缺铁比较明显的情况下才有较大变动。若选用血清铁蛋白作为观察指标,则可敏锐地反映出试验因素的效应。

④特异性:为了更好地揭示研究问题的本质,观察指标还应具备一定的特异性。例如,在诊断糖尿病时,测定血糖的特异性就比测定尿糖的特异性高。

⑤指标的关联性:选用的指标必须与所研究的题目具有本质性联系,且能确切反映试验因素的效应。所选指标是否具有关联性,充分反映了研究者的专业知识与技术水平。

试验效应指标应当同时兼顾其灵敏性和特异性,尽量使灵敏性和特异性都高。

(2)指标的观察。

①对试验效应的观察应避免偏倚。

②处理与效应之间往往存在一定关系。如在药理或毒理试验研究中,药物剂量与效应通常呈“S”形曲线关系,低于某剂量时不引起机体反应;剂量高时,机体反应出现一个峰值;再增大剂量,机体反应不再增加。在两个极值之间存在一个正比关系。因此,应选择一个合适的试验剂量。通常,选择导致半数试验动物产生所期望的剂量,称为半数有效量或半数致死量。

3.2 试验设计的基本原则

内科学科研设计是指对某一项内科学研究的具体内容和方法的安排。如果研究设计存在缺陷(如样本代表性、可比性、混杂等),就可能影响研究结果的可靠性和研究价值。另外,还要考虑研究对象(样本)的数量,样本量过少或过多都会对研究结果造成影响。一个缜密而完善的研究设计,能合理地安排各种研究因素,严格控制各种误差,用较少的人力、物力和时间,最大限度地获得足够而可靠的资料。合理的研究设计是顺利开展各种医学研究的前提条件,也是得到预期结果、达到事半功倍的保证。为了达到上述目的,在内科学试验设计中必须遵循 Ronald A. Fisher 早在 20 世纪 20 年代提出的三项基本原则:随机化原则、对照原则、重复原则。

3.2.1 随机化原则

该原则指被研究的对象样本是从总体中任意抽取的,抽取时每个观察单位都有相等机会被抽取。常用的随机化法有抽签法、抓阄法、扔硬币法、随机数字表法等。随机化的核心是机会均等。随机化是医学研究中一项非常重要的原则,随机化是试验性研究中保证组间均衡、可比的重要手段。在医学研究中,随机化包括两方面的内容:随机化抽样和随机化分组。

(1)随机化抽样。随机化抽样是指总体中的每一个个体都有相同机会被抽中作为研究对象(样本)。在医学研究中,常见的随机化抽样方法有单纯随机化抽样、系统随机化抽样、分层随机化抽样、整群随机化抽样以及多阶段抽样随机化方法。

(2)随机化分组的常用方法。在试验性研究中,除了干预因素外,试验组和对照组在非研究因素的分布上一致,才能消除这些因素对试验结果的影响。通过随机化分组,可以获得有均衡性的试验组和对照组。常用的随机分组方法有完全随机化分组、配对设计、区组随机化分组、分层随机化分组和整群随机化分组等。

3.2.2 对照原则

该原则指在试验研究中,任何一个试验组都必须有与之具有可比性的另一组,也就是说,必须有对比的基础。这个被作为对比基础的组称为“对照组”。在任何一项科学研究中,必须设置“合理对照组”的要求,称为“对照原则”。对照的意义在于通过它消除和减少试验误差。在医学研究中,除了正常解剖生理数据的调查以外,都要有对照。对照的基本要求是,除了对试验因素做有计划的变化外,试验组与对照组的条件应尽量保持一致。

要消除非试验因素的影响,使试验因素的真实效应充分显露出来,必须设立对照。常见的对照形式有空白对照、安慰剂对照、阴性对照、试验对照、阳性对照、标准对照、自身对照、历史对照。

1. 空白对照

空白对照指无干预措施,对照组不加任何处理措施;常用于干预试验疗效研究,以评定测量方法的准确性,观察试验是否处于正常状态,也可排除自然因素或自愈因素对试验结果的影响。如观察某种药物治疗某种传染病的效果,试验组动物使用该药物,对照组动物不使用任何药物,最后比较两组的血清学和流行病学指标。在对皮炎等有自愈倾向的疾病进行防治效果研究时,空白对照

可以很好地说明疾病的痊愈是防治的效果还是自然痊愈。但需要注意的是,空白对照也有它的缺点,即由于不给予任何治疗措施,在一些疾病的治疗试验中可能会违背医学伦理原则。

【例3-5】 某研究者欲通过局部注射阳离子脂质体转染剂和质粒混合物寻找外源性血管内皮生长因子体内转染的有效途径,以及其基因表达对于骨折愈合的促进作用。设计:将40只体重为230~250g的成年雄性SD大鼠随机分为试验组和对照组,每组20只。方法:将40只大鼠建立右侧股骨干骨折模型,于股骨中段切断,断端间相距1mm,于髁间倒置入直径1mm的克氏针,固定骨折断端。试验组将脂质体转染剂100 μ L和pBLAST49-mVEGF 100 μ g质粒混合物两侧骨膜下多点注射,对照组注射等量生理盐水。

【实例辨析】 本例研究者将40只成年雄性SD大鼠随机分为试验组和对照组,试验组将脂质体转染剂100 μ L和pBLAST49-mVEGF 100 μ g质粒混合物两侧骨膜下多点注射,对照组仅注射等量生理盐水,由此可知本例采用的对照形式为空白对照形式。

2. 安慰剂对照

对照组给予安慰剂。安慰剂是一种在外形(剂型、颜色、大小、气味、味道)上与试验药物完全相同,但又不具有特异有效成分的制剂,常用乳糖、淀粉和生理盐水制成。使用安慰剂对照主要是为了避免心理因素对试验结果的影响;也可消除疾病自然进程的影响,观察到试验药物的真正作用。考虑到伦理学原则,安慰剂对照一般用于所研究的疾病尚无有效的防治药物或使用后不会影响到对照组研究对象的健康。同时,应用这一对照,要注意安慰剂效应的影响。目前,安慰剂对照在医学研究中越来越受到重视,但在实际应用中误将安慰剂对照视为空白对照的也不乏其例。

【例3-6】 某研究者欲观察氧化苦参碱胶囊治疗慢性乙型肝炎的疗效及安全性,选择慢性乙型肝炎患者144例,随机分为氧化苦参碱胶囊组72例和空白对照组72例完成52周治疗并随访12周,进行多中心、随机、双盲、安慰剂对照设计的临床试验。治疗前后及停药12周后观察患者临床症状、肝功能、血清乙型肝炎病毒(HBV)标志物和不良反应等。试验药品和方法:氧化苦参碱胶囊由宁夏药物研究所提供(批号:990426)。空白对照药大小、颜色、外形、口味等与氧化苦参碱胶囊一致。氧化苦参碱胶囊组300mg,每日3次,口服52周;空白对照组72例,空白胶囊3粒,每日3次,口服52周。在筛选评价完成后,将符合入选标准的患者分层,随机化设定药物编号,按治疗入选顺序随机分配到氧化苦参碱胶囊组和空白对照组完成52周治疗。

【实例辨析】 本例研究的题目为:氧化苦参碱治疗慢性乙型肝炎的随机双盲对照多中心研究。但作者在描述研究对象分组时提到“本试验共入组144例患者,随机分为氧化苦参碱胶囊组和空白对照组,每组72例”且在描述试验药品时提到“空白对照药大小、颜色、外形、口味等与氧化苦参碱胶囊一致”。由此可知,本例研究所采用的对照形式应为安慰剂对照形式,并非空白对照形式。而本文研究者却在多处提到“空白对照”,故犯了将安慰剂对照和空白对照混淆的错误。

3. 试验对照

为了有效地控制影响试验结果的非试验因素,仅采用空白对照是不够的,此时可以使用试验对照,即对照组的操作条件与试验组一致。例如青霉素过敏试验,以青霉素溶媒为试验对照,可排除由溶媒引起的过敏反应。再如观察某中药预防某种传染病的效果,试验组服用该中药,同时每天进行居住环境的消毒、换气;对照组不服用该中药,但和试验组一样每天进行居住环境的消毒、换气。

【例3-7】 某研究者欲观察专用肠内营养制剂(EN-S)对ANP大鼠蛋白质代谢的调节作用,将32只6~7周龄、体重为200g的清洁级SD健康成年雄性随机分为对照组、EN-S组、EN组和PN组4组,每组8只。将对照组的8只大鼠切开腹腔后,仅轻翻胰腺后缝合切口,后3组先采用胰腺被

膜下多点缓慢均匀注入 3.8% 牛磺胆酸钠 1mL 的方法制备 ANP 大鼠模型, 然后对 EN 组和 EN-S 组待制模后 6h 给予少量生理盐水冲洗肠道, 12h 时开始肠内滴注 EN 或 EN-S 制剂, 起始速滴 1.0mL/h, 逐渐增加至制模后 36h 时滴速达 2.0mL/h, 50mL/d, 并维持这一速度。PN 组大鼠静脉营养液滴注速度在制模后 36h 内从 1.0mL/h 增加至 2.0mL/h, 50mL/d。在制模后 12h、24h、48h 时间点测定并比较四组动物的淀粉酶、血清蛋白等指标的变化情况。

【实例辨析】 在本例研究中对四组试验动物的干预环境均为胰腺, 只是三个试验组与对照组的不同之处在于三个试验组在胰腺被膜下多点缓慢均匀注入 1mL 3.8% 牛磺胆酸钠, 而对照组只是轻翻胰腺后即缝合切口。故本例所采用的对照形式为试验对照形式。

4. 标准对照

在临床试验中, 考虑到要保护对照组人群的健康不受损害, 有时不宜设立安慰剂对照或空白对照。这种情况下, 可以采用目前公认有效的药物或治疗方法作为对照组的措施, 即标准对照。采用标准对照, 可以起到比较的作用, 即消除非试验因素对研究效应的影响。

【例 3-8】 某研究者欲观察中药穴位敷贴对试验性哮喘豚鼠血清中哮喘相关细胞因子肿瘤坏死因子及白细胞介素 II 的影响, 并与地塞米松的作用效应相比较。方法: 选用健康 Hartly 豚鼠 60 只, 随机分为 4 组, 每组各 15 只。采用卵蛋白致敏制作豚鼠支气管哮喘模型, 用中药穴位敷贴治疗(穴位敷贴组)与地塞米松组(标准对照)及模型组(空白对照)和正常对照组比较。(1)穴位敷贴组于造模成功、哮喘发作的次日开始治疗。治疗穴位选择大椎、肺俞、肾俞。用大块胶布将药物敷贴固定于穴位处。每次敷药 6h, 隔日治疗 1 次, 共治疗 7 次。(2)地塞米松组采用腹腔注射地塞米松 0.5mg/kg, 治疗时间与次数同敷贴组。(3)模型组造模成功后, 令其自然恢复。正常对照组不造模, 不给予任何治疗。(4)各组动物治疗结束后, 血清肿瘤坏死因子和白细胞介素 II 水平的检测采用放射免疫分析法。

【实例辨析】 本例研究中选用健康 Hartly 豚鼠 60 只, 随机分为 4 组, 即穴位敷贴组、地塞米松组(标准对照)、模型组(空白对照)和正常对照组。其中, 穴位敷贴组为试验组, 其他三组为不同形式的对照组。而地塞米松组采用腹腔注射地塞米松 0.5mg/kg, 因为地塞米松为治疗哮喘的传统药物, 故该组为标准对照组。模型组造模成功后, 令其自然恢复, 故模型组应为试验对照组, 并非文中提到的“空白对照组”, 而文中所说的“正常对照组”属于空白对照组。

由本例研究可知, 在同一试验中可采用多种不同的对照形式, 但需根据具体的研究目的选择恰当的对照形式。

5. 自身对照

自身对照是指对照和试验在同一研究对象中进行。研究对象在前、后两个阶段, 分别使用两种不同的干预措施, 比较干预的效果, 或者某种方法治疗前后的比较。自身前后对照设计简单, 但其运用前提是, 如果不给这些研究对象以有效的治疗药物, 其效应指标如病情将保持稳定不变。对于自限性疾病如流行性感、甲型病毒性肝炎等, 不宜设置自身对照。自身对照还有一种形式是同一个试验动物的不同部位、器官的比较, 如左、右眼睛及四肢等。

【例 3-9】 题目: 大白鼠急性癫痫模型的 DWI 自身对照分析。试验方法: 试验用鼠为 50 ~ 100g 雌性 SD(Sprague-Dawley) 大白鼠(20 ~ 30 天, $n = 3$) 试验前经腹腔注射 10% 乌拉坦(0.3mL/100g 体重)麻醉施行气管插管术, 连接微型呼吸机(江湾 I 型, 上海), 气量为 3 ~ 4mL/次, 呼吸频率为 60 次/分。通过腹腔埋管注射骨骼肌肉松弛剂(Pavulon), 剂量按 0.5mg/kg 体重, 并每隔 50min 追加相同剂量。把肌肉松弛后的鼠放入仪器进行 MRI 测定, 然后由腹腔埋管注射 0.3% (rag/mL) 印防己毒素 0.2mL 诱发大白鼠癫痫, 20min 后于相同条件下再次进行 MRI 测定. 从而形成自身对照组

和模型组研究体系。在相同试验条件下分别测定同一只鼠在注射印防己毒素前和注射印防己毒素 20ram 后脑部水的扩散系数,作为自身对照。扩散系数的测定在室温(298K)下进行,用加在身体周围的热水瓶保持鼠体温。

【实例辨析】 本例研究以 50 ~ 100g 雄性 SD(Sprague-Dawley)大白鼠为研究对象,并在相同试验条件下分别测定同一只鼠在注射印防己毒素前和注射印防己毒素 20ram 后脑部水的扩散系数,由此可知是比较同一受试对象试验前后的脑部水扩散系数,因而采用的是自身对照形式。

6. 历史对照

用过去研究的结果作为对照称为历史对照。历史对照不是同期对照,由于时间不同,试验条件不同,往往缺乏可比性,一般不建议使用。

【例 3-10】 题目:SD 大鼠碱性磷酸酶活性变异及正常历史对照范围研究。试验方法:以本试验室最近进行的 12 次大鼠亚慢性毒性试验研究为数据来源,所有大鼠均来自于同一供应商的同样品系,饲养条件、研究方法及试验周期均相同。以这 12 次试验的阴性对照组动物为研究对象,分别计算每次试验中不同性别动物在不同时间点(试验中期和试验末期)碱性磷酸酶活性的均值,分析比较不同性别和不同时间点碱性磷酸酶的变化;并计算 12 次试验碱性磷酸酶活性均值的最大值和最小值来确定本试验室 SD 大鼠碱性磷酸酶活性的正常历史对照范围。

【实例辨析】 本例研究以本试验室最近进行的 12 次大鼠亚慢性毒性试验研究为数据来源,计算 12 次试验碱性磷酸酶活性均值的最大值和最小值,以确定本试验室 SD 大鼠碱性磷酸酶活性的正常历史对照范围。由此可知,本研究确定的碱性磷酸酶活性值将作为以后类似研究的参照值,为历史对照形式。

另外,除上述对照形式之外,其他对照形式如交叉对照、潜在对照等对照也被广泛采用。具体内容可参考相关书籍,在此不再详述。

3.2.3 重复原则

重复是保证科研成果可靠性的重要措施之一。重复有两层含义:

(1)指试验过程是多次重复进行的,做到这一点与试验的样本大小有关;样本大,重复的机会多;样本小,重复的机会少。

(2)设计中提出的方法,别人也能重复进行。

重复是指为提高科研的科学性和可靠性,在相同条件下进行多次的观察或多次的研究。广义的重复包括样本数量的重复、观察次数的重复和研究结果的重复。狭义的重复即样本数量的重复。观察次数重复指的是对同一试验对象进行多次观察或测量,以提高观测结果的精确性。一般要求对某项指标至少观测三次。研究结果的重复即重复试验以验证相同条件下结果的重现性,保证结果的可靠性。无法重现的研究是没有科学意义的。样本数量的重复就是对多个试验对象进行独立试验或观察,防止把偶然现象当成必然现象,把个别情况当成普遍情况,甚至错误地推广到总体。在本节中讨论的重复主要指样本含量。

重复的意义在于控制抽样误差,保证结果的可靠性。通过试验对象内部的重复,可以总结总体内部的变异规律,估计抽样误差的大小。

随机化、对照、重复和均衡是医学科研设计中的基本原则,再加上恰当地运用盲法(有利于降低和消除受试者和研究者心理因素对结果造成的影响、提高受试者依从性),是保证研究结果真实可靠的基本原则和重要举措。随机化抽样可以得到一个具有较高客观性和代表性的样本,而随机化分组可以使对比组间,在一切非研究因素方面达到极好的均衡性,即对比组间具有很好的可比性;设立合理的对照组可以排除非研究因素对试验结果的影响;在一项研究中,要求各个小组内应

有足够的样本含量，以便使随机变量的变化规律能充分地显露出来；而均衡性则是一个把关的原则，它要求对比组间在一切非研究因素上处于均衡状态，只有在试验过程中的每一个环节上都做到尽善尽美时，才有可能最终实现均衡性要求。

3.3 内科试验设计方法

内科学科研设计中常要考虑如何进行动物试验的问题，因为很多阐明疾病及疗效机制的试验不可能或不应该在病人身上进行，因此要依赖于动物试验。内科学的动物试验设计方法主要是对试验中的处理因素进行合理的安排，以最小的投入，达到经济、高效的最大收益的目的。

若能找到与人类疾病相同的动物自发性疾病，当然最好。例如，日本的大白鼠原发性高血压就是研究人类原发性高血压的理想模型，猪自发性冠状动脉粥样硬化是研究人类冠心病的理想模型；自发性狗类风湿性关节炎与人类幼年型类风湿性关节炎十分相似，也是一种理想模型，等等。与人类完全相同的动物自发性疾病模型毕竟不可多得，往往需要人工制造。

为了尽量做到与人类疾病相似，首先要注意试验动物的选择。例如，小鸡最适宜做高血脂症的模型，因为它的血浆甘油三脂、胆固醇以及游离脂肪酸水平与人十分相似，低密度和极低密度脂蛋白的脂质构成也与人相似。其次，为了尽可能做到试验结果的逼真性，要在实践中对方法不断加以改进。例如结扎兔阑尾血管，固然可能使阑尾坏死穿孔并导致腹膜炎，但这与人类急性梗阻性阑尾炎合并穿孔和腹膜不一样，如果给兔结扎阑尾基部而保留原来的血液供应，由此而引起的阑尾穿孔及腹膜炎就与人的情况相似，因而是一种比较理想的方法。

如果动物试验的情形与临床情况不相似，在动物身上有效的治疗方案就不一定能用于临床，反之亦然。例如，动物内毒素性休克(Endotoxin Shock，单纯给动物静脉输入细菌及其毒素所致的休克)与临床感染性(脓毒性)休克(Septic Shock)就不完全一样，因此对动物内毒素性休克有效的疗法长期以来不能被临床医生所采用。现在，有人改向结扎胆囊动脉和胆管的动物胆囊中注入细菌，模拟人类感染性休克的模型，认为这样动物既有感染又有内毒素中毒，就与临床感染性休克相似。

为了判定所进行的动物试验是否与人相似，需要进行一系列的检查。例如有人检查了动脉压、脉率、静脉压、呼吸频率、动脉血 pH、动脉氧分压和二氧化碳分压、静脉血乳酸盐浓度以及血容量等指标，发现一次定量放血法造成的休克模型与临床出血性休克十分相似，因此认为这些方法进行的动物试验是一种较理想的试验。同理，按中医理论用大黄喂小鼠使其出现类似人的“脾虚症”，如果又按中医理论用四君子汤把它治好，那么就有理由把它看成人类“脾虚症”的动物模型。

理想的动物试验应该是可重复的，甚至是可以标准化的。例如用一次定量放血法可百分之百地造成出血性休克，百分之百的死亡，这就符合可重复性和达到了标准化要求。又如用狗做心肌梗死模型照理很合适，因为它的冠状动脉循环与人相似，而且在试验动物中它最适宜做暴露心脏的剖胸手术，但狗结扎冠状动脉的后果差异太大，不同狗的同一动脉、同一部位的结扎，其后果很不一致，无法预测，无法标准化。相反，小白鼠、地鼠和豚鼠结扎冠脉的后果就比较稳定一致，可以预测，因而可以标准化。

为了增强动物试验时的可重复性，必须在动物品种、品系、年龄、性别、体重、健康情况、饲养管理；试验及环境条件，季节、昼夜节律、应激、室温、湿度、气压、消毒灭菌；试验方法步骤；药品生产厂家、批号、纯度规格、给药剂型、剂量、途径、方法；麻醉、镇静、镇痛等用药情况；仪器型号、灵敏度、精确度；试验者操作技术熟练程度等方面保持一致，因为一致性是重复性的可靠保证。复制的动物模型应该力求可靠地反映人类疾病，即可特异地、可靠地反映某种疾病或某种机

能、代谢、结构变化,应具备该种疾病的主要症状和体征,经化验或X光照片、心电图、病理切片等证实。若易自发地出现某些相应病变的动物,就不应加以选用,易产生与复制疾病相混淆的疾病者也不宜选用。例如铅中毒可用大白鼠做模型,但有缺点,因为它本身容易患动物地方性肺炎及进行性肾病,后者容易与铅中毒所致的肾病相混淆,不易确定该肾病是由铅中毒所致还是由它本身的疾病所致。用蒙古沙土鼠就比较容易确定,因为一般只有铅中毒才会使它出现相应的肾病变。

对于供医学试验研究用的动物模型,在复制时,应尽量考虑到今后临床应用和便于控制其疾病的发展,以利于研究的开展。如雌激素能终止大鼠和小鼠的早期妊娠,但不能终止人的妊娠。因此,选用雌激素复制大鼠和小鼠终止早期妊娠的模型是不适用的,因为在大鼠和小鼠筛选带有雌激素活性的药物时,常常会发现这些药物能终止妊娠,似乎可能是有效的避孕药,但一旦用于人则并不成功。所以,如果知道一个化合物具有雌激素活性,用这个化合物在大鼠或小鼠上观察终止妊娠的作用是没有意义的。又如选用大小鼠做试验性腹膜炎就不适用,因为它们对革兰氏阴性细菌具有较高的抵抗力,很不容易造成腹膜炎。有的动物对某致病因子特别敏感,极易死亡,也不适用。如狗腹腔注射粪便滤液引起腹膜炎,很快死亡(80%会在24h内死亡),来不及做试验治疗观察,而且粪便剂量及细菌菌株不好控制,因此不能准确重复试验结果。

在进行动物试验时,所采用的方法应尽量做到容易执行和合乎经济原则。灵长类动物与人最近似,复制的疾病模型相似性好,但稀少昂贵,即使猕猴也不可多得,更不用说猩猩、长臂猿。幸好很多小动物(如大小鼠、地鼠、豚鼠等)也可以复制出十分近似的人类疾病模型。它们容易做到遗传背景明确,体内微生物可加控制、模型性显著且稳定,年龄、性别、体重等可任意选择,而且廉价易得、便于饲养管理,因此可尽量采用。除非不得已或除一些特殊疾病(如痢疾、脊髓灰白质炎等)研究需要外,尽量不用灵长类动物。除了在动物选择上要考虑易行性和经济性原则外,而且在模型复制的方法上、指标的观察上也都要注意这一原则。

动物试验时必须强调从研究目的出发,熟悉诱发条件、宿主特征、疾病表现和发病机理,即充分了解所需动物模型的全部信息,分析能否得到预期的结果。例如诱发动脉粥样硬化时,草食类动物兔需要的胆固醇剂量比人高得多,而且病变部位并不出现在主动脉弓。病理表现为以纤维组织和平滑肌增生为主,可有大量泡沫样细胞形成斑块,这与人类的情况差距较大。因此要求研究者懂得,各种动物所需的诱发剂量、宿主年龄、性别和遗传性状等对试验的影响,以及动物疾病在组织学、生化学、病理学等方面与人类疾病之间的差异。要避免选用与人类对应器官相似性很小的动物疾病作为模型材料。为了增加所复制动物疾病模型与人类疾病的相似性,应尽量选用各种敏感动物与人类疾病相应的动物模型。

动物试验应适合多数研究者使用,容易复制,试验中便于操作和采集各种标本。同时应该首选一般饲养员较熟悉而便于饲养的动物作为研究对象。这样,就不需特殊的饲养设施和转运条件,经济上和技术上容易得到保证。

此外,动物来源必须充足,选用多胎分娩的动物对扩大样本和重复试验是有益的。尤其对慢性疾病模型来说,动物须有一定的生存期,便于长期观察使用,以免模型完成时动物已濒于死亡或死于并发症。

在自然环境中观察野生动物有助于正确评价自然发病率和死亡率。但记录困难,在试验条件下维持有一定难度,且对人和家畜有直接和间接的威胁,使用时要特别加以注意。因此,复制模型时必须注意动物种群的选择,要了解各类动物种群的特点和对复制动物的影响。

用于生物医学研究的动物种群,可按其遗传成分和其环境被研究人员控制的程度,分为三种基本类型。

(1)实验室类型。它们可提供最大程度的遗传和环境操作。

(2)家养类型。不论是乡村或城市饲养的,人类对其干扰的程度不同,且动物环境与人类环境可能极为接近。

(3)自然生态类型。几乎没有人造的干扰。

可能对某种动物(啮齿目、食肉目、兔形目)可按所有三类类型进行研究,这就增加了对环境和遗传因素做比较研究的可能性。在选用三类动物种群复制动物模型时,必须了解它们各自的优点和缺点。

动物试验的成败往往与环境的改变有密切关系。拥挤、饮食改变、过度光照、噪声、屏障系统的破坏等中的任何一项被忽视都可能给模型动物带来严重影响。除此以外,复制过程中固定、出血、麻醉、手术、药物和并发症等处理不当,同样会产生难以估量的恶果。因此,要求尽可能使模型动物处于最小的变动和最少的干扰之中。

不能盲目地使用近交系动物,不然会导致不能控制的因素进入试验。例如自发性糖尿病大鼠(BB、Wistar)除具有糖尿病临床特征外,还发现多种病理变化(外周神经系统严重病变、睾丸萎缩、甲状腺炎、胃溃疡、恶性淋巴瘤等)。因此要有目的地选择。半个世纪以来,近交系的开发不断提供着新的动物模型材料,大、小鼠疾病作为模型在医学使用量已高达70%~90%。

动物进化的高级程度并不意味着所有器官和功能接近于人的程度。进行动物试验时,在条件允许的情况下,应尽量考虑选用与人相似、进化程度高的动物作模型。但不能因此就认为进化程度越高的动物的所有器官和功能越接近于人。例如,非人灵长类诱发动脉粥样硬化时,病变部位经常在小动脉、即使出现在大动脉也与人类分布不同。据报道用鸽子作这类试验时,胸主动脉出现的黄斑面积可达10%,镜下变化与人也比较相似,因此也被研究者广泛使用。

正确地评估动物疾病模型,应该懂得没有一种动物模型能完全复制人类疾病的真实情况,动物毕竟不是人体的缩影。模型试验只是一种间接性研究,只可能在一个局部或几个方面与人类疾病相似。因此,模型试验结论的正确性只是相对的,最终必须在人体身上得到验证。复制过程中一旦出现与人类疾病不同的情况,必须分析其分歧范围和程度,找到相平行的共同点,正确评估哪些是有价值的。

除此之外,对于内科学动物试验而言,总的来说应该是两种动物、三个剂量,即至少应在两种动物体上进行,兼顾雌雄性别。剂量至少分为高、中、低三个档次,才能充分反应药物的毒性。同时设一溶剂(或赋形剂)对照组或已知药物毒性对照组。试验动物应该是年轻健康、来源清楚的。最为理想的是选用与人类代谢相似、对药物敏感的动物,在数量上应能满足试验终止后统计学处理的要求,并备有部分动物供停药后观察其毒性反应的可逆性。给药途径应与临床拟用途径相一致。观察指标应能反映试验动物功能及结构的完整性是否受到损害。对于一般毒性症状应定期细致地进行观察,必要时进行中期血液生化检查,但中期检查不必过于频繁,以免影响试验动物的正常活动规律。

3.3.1 剂量设计的方法

在进行动物试验时,选择动物要注意动物的种类及品系、起始月(周)龄、性别及数量、饲养管理、给药期限、剂量等,千万不能疏忽大意,否则试验难以得到可信的结果,严重时甚至使试验不能进行到底。其中,剂量选择是首先遇到的问题。剂量设计的方法有多种。

(1)半数致死法。根据药物急性毒性的LD值决定3个剂量。在一般大鼠3个月长期毒性试验中,高、中、低3个剂量可分别采用LD₅₀的1/10、1/50和1/100剂量;狗可用大鼠试验用的一半左右的剂量。此法比较粗略,差异大,一般只作为预测的参考。若急性毒性和长期毒性试验用的是同一种动物,则参考价值可大些;反之,参考价值不大。

(2)短期蓄积试验过度法。在获得急性毒性的有关资料后,应根据所提供的有关信息、入毒性的量-效关系、毒性出现的剂量范围及敏感动物等,先过渡到7~21天的短期蓄积性试验,初步将急性毒性的参数外推于长期毒性试验。

(3)药代动力学法(MBS法)。根据药代动力学的结果,参照最大有效安全浓度(MCD),如大鼠3个月长期毒性试验时,低、中、高3个剂量可分别采用对大鼠的1.0MCD、3.3~6.6MCD和10~20MCD,狗一般可采用大鼠剂量的一半左右。此法有参考价值,但也要注意动物的异同性,尤其是用小鼠的药代动力学结果时更应慎重。此外,有不少药代动力学和长期毒性是同期进行的,甚至是事后补做的,故国内实际用此法的不多。

(4)最大耐受量法(MTD法)。用大鼠急性毒性的最大无症状剂量,即一个最大耐受量(1MTD)、1/3MTD和1/10MTD分别为大鼠长期毒性的3个剂量,对狗和猴可考虑用大鼠剂量的一半左右,或测出单次给药狗或猴的MTD,余同大鼠。此法对大部分动物有参考价值,但少数蓄积性大的则不合适。

(5)拟用临床剂全法(ACD法)。根据同类型药物或国外资料或拟推荐临床剂量,一般3个月长期毒性试验时,大鼠可采用50~100倍、25~50倍、10~20倍临床剂量,狗可分别采用30~50倍、15~25倍、5~10倍,猴也可用略低于狗的剂量进行。此法参考价值较大,也较常用。

(6)等效剂量法。根据人和动物的等效剂量比值,再按一定倍数扩大。此法参考价值较大,尤其是对抗肿瘤药、镇痛药和外用药等更有意义。

上述方法是前人在一定条件下从对某些药物或化合物的毒性研究中总结出的经验,有参考价值,尤其是后3种方法,受到毒理学界的重视。但必须认识到某些药物或化合物千差万别,受试药各种各样,任何机械照抄照搬都是不可取的,而应当是个案研究,对具体药物进行认真而细心的预试,这才是摸准剂量的关键。

3.3.2 试验设计方法

从量化的角度对动物试验进行设计时,常见的试验设计方法有配对设计(matched—pairs design)、交叉试验设计(cross—over design)、单因素 $K(K \geq 2)$ 水平设计、随机区组设计(randomized block design)、平衡不完全随机区组设计(balanced incomplete blocks design)、拉丁方设计(Latin—Square design)、析因设计(factorial design)、正交设计(orthogonal design)、序贯试验设计(sequential trial)等。

1. 配对设计

配对设计是将受试对象配成对子,随机给予每对中的两个体以不同处理。配对条件为主要的非试验因素。在动物试验中,常将种属、性别、年龄、体重相近的两动物配成对子。

配对设计的形式:自身配对同一对象接受两种处理,如同一标本用两种方法进行检验,同一患者接受两种处理方法。异体配对将条件相近的试验对象配对,并分别给予两种处理。

配对设计的特点:优点是,比较理想地控制了非试验因素的影响,均衡性较好,减少抽样误差。缺点是,观察对象要经过挑选,特别是临床试验中病例较少时,样本含量较少。

配对设计的统计分析:当各对定量数据的差量服从正态分布时,用配对比较的 t 检验;当不服从正态分布或分布未知时,用配对符号秩和检验。

2. 交叉试验设计

将A、B两种处理先后施于同一批试验对象,随机地使半数对象先接受A,后接受B;另一半对象先接受B,后接受A。两种处理在全部试验过程中“交叉”进行,称为交叉试验。由于A和B处于先后两个试验阶段的机会是相等的,因此平衡了试验顺序的影响,而且能把处理方法之间的差别与

时间先后之间的差别分开来分析。

确定受试对象的例数必须是偶数，并编号，尽量使相邻的第 1、2 号非试验因素近似，第 3、4 号非试验因素近似，余类推。随机地确定各单号的接受两种处理的顺序，而各双号的顺序，与其前一个单号的顺序相反。

统计分析方法：当试验因素对应的定量数据的差量服从正态分布时，用交叉设计定量资料方差分析；当其不服从正态分布时，用秩和检验。

交叉试验设计的特点：优点是节省样本数(特指单个型交叉设计，即两种处理先后作用与同一个受试对象)，且均衡性较好；缺点是易于导致接受第二次处理后的试验效应被第一次处理所“污染”。故特别指出，在下列三种场合下不宜采用单个体型交叉设计：两次处理的时间间隔不能过长，如比较两种抗菌素治疗肺炎的效果，病人一旦痊愈就不能再用另一种抗生素；当药物有蓄积作用或排泄缓慢，使得第二次处理前不能彻底排除第一次处理的干扰；当第一次处理对指标取值的影响是持久的，即指标的数值不可能在施加第二次处理前恢复原先的水平，也不适用单个体型交叉设计。

【例 3-11】 试验两种不同配方的减肥药物 a1 和 a2，采用交叉设计将 20 只患有肥胖症的大鼠随机分为甲、乙两组，甲组前 4 周服 a1 药，后 4 周服 a2 药；乙组前 4 周服 a2 药，后 4 周服 a1 药。试验结果如表 3-4 所示(应先按体重相近，将 20 只受试对象配成 10 对)。

表 3-4 受试者的体重下降值(kg)

给 药 顺 序	受 试 验 者	前 4 周	后 4 周
甲组 a1 → a2	1. 1(1)	6. 129	-0. 454
	2. 1(2)	2. 497	0. 908
	⋮	⋮	⋮
	9. 2(9)	1. 498	1. 135
	10. 1(10)	8. 172	4. 313
乙组 a2 → a1	1. 2(11)	4. 449	2. 043
	2. 2(12)	4. 994	1. 816
	⋮	⋮	⋮
	9. 1(19)	1. 589	1. 271
	10. 2(20)	0. 136	1. 589

说明：本例中若能取 10 窝大鼠，每窝两只性别相同、体重接近，配成一对，采用多个型交叉设计，结果的准确性会得到很大提高。

3. 单因素 $K(\geq 2)$ 水平设计

单因素 $K(\geq 2)$ 水平设计可分为两种情形。其一，试验因素与受试对象无直接联系(如试验因素为药物种类、药物剂量、作用时间等)，需要经过两次随机，才能将受试对象分配到各组中去。第一次，从特定的总体中随机抽取规定数目的受试对象；第二次，将抽取的受试对象随机地分配到各组中去。其二，试验因素与受试对象有直接联系(如试验因素为属性因素，通常为性别、血型、年龄分组、病情、病程等)，只需从属性因素各自的水平组中随机抽取规定数目的受试对象，不需进一步分组。各组受试对象例数可以相等，也可以不相等。

单因素 $K(\geq 2)$ 水平设计的特点：优点是设计和统计分析比较简单，缺点是试验效率较低且只能分析一个因素。

单因素 $K(\geq 2)$ 水平设计的统计分析方法如下。

(1) 两处理组(即 $K=2$)。

①大样本(当 n_1 及 n_2 均大于 100 时)当资料为定量资料时,可用 Z 检验(以标准正态分布为其理论根据);当资料为定性资料时,可用 χ^2 检验或 Fisher 精确检验。

②小样本定量资料:当服从正态分布且满足方差齐性时,采用 t 检验;当服从正态分布但不满足方差齐性时,采用 t' 检验;当不服从正态分布或分布未知时,采用秩和检验;当资料为定性资料时可用 Fisher 精确检验。

(2)多个处理组(即 $K > 2$)。

为定量资料时,当服从正态分布且满足方差齐性时,采用单因素 $K(\geq 2)$ 水平设计一元定量资料方差分析;当不服从正态分布或分布未知时,采用相应设计一元定量资料秩和检验;当资料为定性资料时采用 χ^2 检验或 Fisher 精确检验。

【例 3-12】 研究髂静脉压迫综合征及其他原因所致的静脉压迫性损伤的病理机制。成年 SD 大鼠 48 只,雌雄不限,体重为 $300\text{g} \pm 10\%$ 。将 48 只大鼠随机分为 4 组,3% 水合氯醛(按 $1\text{mL}/100\text{g}$ 体重)腹腔内注射麻醉,0.5% 硫化钠腹部脱毛,左下腹沿股静脉方向做斜行切口。暴露股静脉并沿股静脉向上分离至后腹膜壁层后方显露左髂总静脉,其中 3 组用 4.5#、5#、6#等型号注射针头小心放置于左髂总静脉旁,用相同的拉力将左髂总静脉和左髂总动脉与针头一同结扎,缓慢抽出针头,即可造成髂总静脉不同程度的受压、狭窄,同时,结扎腹壁下静脉等可视的静脉属支。余下一组手术方法完全相同但不结扎血管,为对照组。术后正常喂养。

【实例辨析】 48 只大鼠随机分为 4 组,接受不同处理,为完全随机组设计。

4. 随机区组设计

随机区组设计是配对设计的扩大。当处理组数为三个或三个以上时,将非试验因素相同或相似的受试对象组成一个区组,每个区组内受试对象的个数是试验因素水平数的整数倍(通常就等于其水平数)且被随机地均分入各处理组中去。

(1)随机区组设计的特点:各个处理组中的受试对象不仅数量相同,而且比较均衡。这种试验设计,既减小了抽样误差,又可以分析出处理组(即试验因素)及随机区组(即重要非试验因素)两个因素的影响。其缺点与配对设计相同,不易找全找准重要的非试验因素来构造复合型的“区组因素”。

(2)缺项估计:随机区组试验的特点是各处理组的受试对象数相等,各区组的受试对象数也相等。例如在试验过程中,因故造成某个数据丢失,如试验过程中死去一只动物或出现仪器故障,此时如果将此区组整个去掉,对信息是一种损失。但由于缺少一个数据,无法进行随机区组设计一元定量资料方差分析。可以采用缺项估计加以解决。

【例 3-13】 研究阿魏酸与芍药甙配伍对神经元的存活和生长的影响。取出生 24h 内的新生 SD 大鼠进行常规碘酒、酒精消毒后,在无菌操作下剪去头皮、颅骨,取出大鼠的脑,放入含有 D-Hank's 液的大培养皿中。在解剖显微镜下,分离出海马和隔区组织,分别置于盛有 DHank's 液的青霉素瓶中,清洗 3 次,然后用眼科剪将组织剪碎,吸去 D-Hank's 液加入 0.125% 的胰蛋白酶,在 37°C 的温箱中消化 10~15min 后用含有 15% 血清的 DMEM/F12 培养基终止消失。用细口吸管吹打细胞,将细胞悬液移入离心管中,以 $1000\text{rpm}/\text{min}$ 离心 10min,弃去上清液,加入 DMEM/F12 液用 8 号针头的注射器吹打成均匀一致的细胞悬液。将细胞移入青霉素瓶中,再经 $0.22\mu\text{m}$ 不锈钢滤网过滤后,用 0.14% 台盼蓝染色法计数细胞活力,按 2×10^5 个细胞/毫升的密度接种在涂有多聚赖氨酸的 96 孔培养板内,每孔加入 $50\mu\text{L}$ 细胞悬液,放入 CO_2 培养箱培养并观察细胞生长情况。

培养 24h 后,分为 4 个组,即阿魏酸组、芍药甙组、配伍组、对照组。其中,阿魏酸组加入阿魏酶终浓度为 $25\mu\text{g}/\text{mL}$ 。芍药甙组加入芍药甙终浓度为 $715\mu\text{g}/\text{mL}$ 。配伍组加入阿魏酸和芍药甙终浓

度 $25\mu\text{g/mL}$ 加 $715\mu\text{g/mL}$ 。对照组以培养基替代。继续培养 3d，在倒置相差显微镜下观察神经元的存活生长情况。

【实例辨析】 本例中的“配伍”是中医药中的一个专有名词，即药物有单用与配伍使用(就是联合用药)两种，千万不要与医学统计学上所讲的“随机区组设计(简称为配伍组设计)”相混淆。前述所描述的试验涉及两个因素：一个因素为“是否用阿魏酸”，而另一个因素为“是否用芍药甙”。当这两个因素对某定量指标的影响有主次之分时，其设计类型属于两因素嵌套设计；反之，属于两因素析因设计。

5. 平衡不完全随机区组设计

一般每个区组内受试对象个数 k 应等于处理组数 v ，但有时处理组数 v 多于每个区组所能容纳的受试对象个数，即 $v > k$ 。此时每个区组不能把所有的处理都安排进去，可采用平衡不完全随机区组设计，简称 *BIB* 设计。有现成的表格可供查阅并使用，此处从略。

6. 拉丁方设计

用 r 个拉丁字母排成 r 行 r 列的方阵，使每行及每列中每个字母都只出现一次，这样的方阵称为 r 阶拉丁方，或 $r \times r$ 拉丁方。

拉丁方设计使用与统计分析拉丁方试验设计用于三个因素(一个试验因素、两个重要非试验因素)，三个因素的水平数相同，且因素间没有交互作用。优点：试验次数减少到最小，并且均衡。

统计分析：拉丁方设计一元定量资料方差分析。

【例 3-14】 为了研究 5 种不同温度对蛋鸡产蛋量的影响，将 5 栋鸡舍的温度设为 A、B、C、D、E，把各栋鸡舍的鸡群的产蛋期分为 5 期，由于各鸡群和产蛋期的不同对产蛋量有较大的影响，因此采用拉丁方设计。把鸡群和产蛋期作为单位组设置，以便控制这两个方面的系统误差。

【实例辨析】 总共有 3 个因素，每个因素有 5 个水平，适合做拉丁方设计。

拉丁方设计步骤如下。

(1) 选择拉丁方时应根据试验的处理数和横行、纵列单位组数先确定采用几阶拉丁方，再选择标准型拉丁方或非标准型拉丁方。此例因试验因素为温度，其水平数为 5；将鸡群作为纵列单位组因素，其水平数为 5；将产蛋期作为横行单位组因素，其水平数也为 5，即试验因素水平数、纵列单位组数、横行单位组数均为 5，则应选取 5×5 阶拉丁方。本例选取前面列出的第 2 个 5×5 标准型拉丁方，即

A	B	C	D	E
B	A	D	E	C
C	E	B	A	D
D	C	E	B	A
E	D	A	C	B

(2) 随机排列。在选定拉丁方之后，如是非标准型时，则可直接按拉丁方中的字母安排试验方案。若是标准型拉丁方，还应按下列要求对横行、纵列和试验处理的顺序进行随机排列。

3×3 标准型拉丁方：纵列随机排列，再将第二和第三横行随机排列。

4×4 标准型拉丁方：随机选择 4 个标准型拉丁方中的一个，然后再将横行、纵列及处理都随机排列。

下面对选定的 5×5 标准型拉丁方进行随机排列。先从随机数字表(I)第 22 行、第 8 列 97 开始，向右连续抄录 3 个 5 位数，抄录时舍去“0”、“6 以上的数”和重复出现的数，抄录的 3 个 5 位数

字为 13542、41523、34521。然后，将上面选定的 5 × 5 拉丁方的纵列、横行及处理按这 3 个 5 位数的顺序重新随机排列。

- ①纵列随机：将拉丁方的各纵列顺序按 13542 顺序重排。
- ②横行随机：再将纵列重排后的拉丁方的各横行按 41523 顺序重排。排列结果如表 3-5 所示。
- ③把 5 种不同温度按第 3 个 5 位数 34521 顺序排列，即 A = 3，B = 4，C = 5，D = 2，E = 1，也就是说，在拉丁方中的 A 表示第 3 种温度，B 表示第 4 种温度等，依此类推。从而得出 5 × 5 拉丁方设计，如表 3-6 所示。

表 3-5 拉丁方设计步骤

选择拉丁方					纵 列 随 机					横 行 随 机				
1	2	3	4	5		1	3	5	4	2				
A	B	C	D	E	1	A	C	E	D	B	4	D	E	A
B	A	D	E	C	2	B	D	C	E	A	1	A	C	E
C	E	B	A	D	3	C	B	D	A	E	5	E	A	B
D	C	E	B	A	4	D	E	A	B	C	2	B	D	C
E	D	A	C	B	5	E	A	B	C	D	3	C	B	D

表 3-6 5 种不同温度对鸡产蛋量影响的拉丁方设计

产 蛋 期	温度的编号				
	鸡群：一	二	三	四	五
I	D(2)	E(1)	A(3)	B(4)	C(5)
II	A(3)	C(5)	E(1)	D(2)	B(4)
III	E(1)	A(3)	B(4)	C(5)	D(2)
IV	B(4)	D(2)	C(5)	E(1)	A(3)
V	C(5)	B(4)	D(2)	A(3)	E(1)

注：括号内的数字表示温度的编号。

由表 3-6 可以看出，第一鸡群在第 I 个产蛋期用第 2 种温度，第二鸡群在第 I 个产蛋期用第 1 种温度，等等。表 3-7 是试验结果。

表 3-7 5 种不同温度对母鸡产蛋量影响的试验结果(单位：个)

产 蛋 期	第一鸡群	第二鸡群	第三鸡群	第四鸡群	第五鸡群	横行和
I	D(23)	E(21)	A(24)	B(21)	C(19)	108
II	A(22)	C(20)	E(20)	D(21)	B(22)	105
III	E(20)	A(25)	B(26)	C(22)	D(23)	116
IV	B(25)	D(22)	C(25)	E(21)	A(23)	116
V	C(19)	B(20)	D(24)	A(22)	E(19)	104
纵列和	109	108	119	107	106	总合计 = 549

注：括号内数字为产蛋量。

对表 3-8 资料进行方差分析后得到如表 3-9 所示的方差分析表。

表 3-8 各种温度(处理)的合计

温 度	A	B	C	D	E
$x_{(k)}$	116	114	105	113	101
$\bar{x}_{(k)}$	23.2	22.8	21.0	22.6	20.2

表 3-9 资料的方差分析表

变 异 来 源	SS	df	MS	<i>F</i>	<i>P</i>
横行间	27.36	4	6.84	4.56	<0.05
纵列间	22.16	4	5.54	3.69	<0.05
温度间	33.36	4	8.34	5.56	<0.05
误差	18.08	12	1.50		
总变异	100.96	24			

经 *F* 检验，产蛋期间和鸡群间、温度间差异有统计学意义。因在拉丁方设计中，横行、纵列单位组因素是为了控制和降低试验误差而设置的非试验因素，所以即使有统计学意义，一般也不对单位组间进行多重比较。而对主要的处理因素，则可以进行多重比较。表 3-10 是对不同温度平均产蛋量间作进行多重比较的情况。

表 3-10 不同温度平均产蛋量多重比较表(q 法)

温 度	平均数 $\bar{x}_{(k)}$	$\bar{x}_{(k)} - 20.2$	$\bar{x}_{(k)} - 21$	$\bar{x}_{(k)} - 22.6$	$\bar{x}_{(k)} - 22.8$
A	23.2	3.0	2.2	0.6	0.4
B	22.8	2.6	1.8	0.2	
C	22.6	2.4	1.6		
D	21.0	0.8			
E	20.2				

温度平均数标准误为：

$$S_{\bar{x}} = \sqrt{MS_e/n} = \sqrt{1.5/5} = 0.55$$

由 $df_e = 12$ 和 $k = 2, 3, 4, 5$ 从 *q* 值表查得临界 *q* 值： $q_{0.05}$ 和 $q_{0.01}$ ，并与 $S_{\bar{x}}$ 相乘得 LSR_a 值（注： LSR 为最小显著极差），列于表 3-11 中。

表 3-11 *q* 值和 LSR 值表

df_e	<i>k</i>	$q_{0.05}$	$q_{0.01}$	$LSR_{0.05}$	$LSR_{0.01}$
12	2	3.08	4.32	1.69	2.38
	3	3.77	5.04	2.07	2.77
	4	4.20	5.50	2.31	3.03
	5	4.51	5.84	2.48	3.21

多重比较结果表明：温度 A、B、D 平均产蛋量明显地高于 E，即第 3、4、2 种温度的平均产蛋量均高于第 1 种温度的平均产蛋量，其余之间差异无统计学意义。第 1 种和第 5 种温度平均产蛋量最低。

7. 析因设计

析因设计是多因素的交叉分组设计，即各个因素的所有水平相互组合在一起，形成多个交叉分组。交叉组数是各因素水平数的乘积。析因设计不仅可以对每个因数各水平间进行比较，还可以分析因素间的交互作用。

交互作用:因素间如存在交互作用，表示各因素不是各自独立的，而是一个因素的水平有改变时，其他因素的效应也随之改变，反之，如不存在交互作用，则表示各因素具有独立性，一个因素的水平有改变时，其他因素的效应不受影响。

【例 3-15】 对 12 例缺铁性贫血小鼠的疗效观察，分为 4 组，分别在一般疗法基础上联合甲、乙两药进行治疗，1 个月后检查各组动物的红细胞数($10^5/mm^3$)，如表 3-12 所示。

第一组采用一般疗法，第二组采用一般疗法 + 甲药，第三组采用一般疗法 + 乙药，第四组采用一般疗法 + 甲药 + 乙药，治疗后的结果如表 3-13 所示。

此题采用析因设计的方差分析后，各组之间的差别有统计学意义(关于析因设计的分析过程，参考第 4 章 4.2 节)，进一步分析交互效应如表 3-14 所示。由此表可以看出，甲、乙两药有协同作用。

表 3-13 缺铁性贫血治疗 1 个月后红细胞增加数

病 例 编 号	第 一 组	第 二 组	第 三 组	第 四 组	合 计
1	0.8	1.3	0.9	2.1	5.1
2	0.9	1.2	1.1	2.2	5.4
3	0.9	1.1	1.0	2.0	4.8
合计	2.4	3.6	3.0	6.3	15.3
均数	0.8	1.2	1.0	2.1	5.1

表 3-14 甲药与乙药交互作用分析表

	不 用 甲 药	使 用 甲 药	合 计
不用乙药	2.4	3.0	5.4
使用乙药	3.6	6.3	9.9
合计	6.0	9.3	15.3

8. 正交设计

正交设计是一种高效、快速的多因素试验方法。它是利用一套规格化的正交表，使每次试验的因素及水平得到合理安排，通过试验结果的分析，获得有用的信息。除了分析主因素外，还可分析部分交互作用的效应。

【例 3-16】 用雌雄大白鼠做试验，通过观察细胞色素 b5 (nmol/mg) 了解正氟醚的毒性作用，同时了解生理盐水和戊巴比妥作为诱导剂对不同性别大白鼠的作用有无不同。试做正交设计。

分析：优于本研究中有三个因素，每个因素有两个水平，现要看各个因素的单独作用和两因素间的交互作用，可采用的方案是析因和正交设计，此处采用正交设计。主要过程如下。

(1) 确定试验因素。有诱导剂(A)、正氟醚(B)、鼠性别(C)的主效应、交互效应 AB、AC 共 5 项。

(2) 选择因素水平，制定因素水平表如表 3-15 所示。

表 3-15 试验因素及水平

水 平	诱导剂 (A)	正氟醚 (B)	大白鼠性别 (C)
1	生理盐水	不用正氟醚	雄性大白鼠
2	戊巴比妥	用正氟醚	雌性大白鼠

(3) 选择合适的正交表并制定试验方案。

根据本题的要求，选择 $L_8(2^7)$ 正交表如表 3-16(关于正交表的选择参考第 4 章 4.5 节)所示。

(4) 表头设计。

表头设计的主要目的是，在正交表中安排试验因素和交互效应，每一个正交表都对应着一个交互效应表，用于说明交互效应所在的列。每一列可有 3 种安排，即安排某个因素、安排交互效应或者空列。表 3-17 是 $L_8(2^7)$ 正交表的交互效应表。在该表中，行的列号和列的列号的交叉点数字为

该两列因素的交互效应列。表 3-17 中第 1 列和第 2 列的交互效应在第 3 列，第 1 列和第 3 列的交互效应在第 2 列，第 6 列和第 7 列的交互效应在第 1 列等。

表 3-16 $L_8(2^7)$ 正交表

处理编号	各列水平代码						
	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

表 3-17 $L_8(2^7)$ 正交表的交互效应表

列 号	交互作用所对应的列号						
	1	2	3	4	5	6	7
1		3	2	5	4	7	6
2			1	6	7	4	5
3				7	6	5	4
4					1	2	3
5						3	2
6							1

本例的正交试验的表头设计如表 3-18 所示。

表 3-18 例 3-17 正交试验的表头设计

因素	1(A)	2(B)	3(AB)	4(C)	5(AC)	6	7
----	------	------	-------	------	-------	---	---

(5) 确定试验方案并记录试验结果如表 3-19 所示。

表 3-19 正交试验结果

处 理	1(A)	2(B)	3(AB)	4(C)	5(AC)	6	7	细胞色素 b5
1	1	1	1	1	1	1	1	0.66(A ₁ B ₁ C ₁)
2	1	1	1	2	2	2	2	0.23(A ₁ B ₁ C ₂)
3	1	2	2	1	1	2	2	0.11(A ₁ B ₂ C ₁)
4	1	2	2	2	2	1	1	0.13(A ₁ B ₂ C ₂)
5	2	1	2	1	2	1	2	0.64(A ₂ B ₁ C ₁)
6	2	1	2	2	1	2	1	0.95(A ₂ B ₁ C ₂)
7	2	2	1	1	2	2	1	0.15(A ₂ B ₂ C ₁)
8	2	2	1	2	1	1	2	0.12(A ₂ B ₂ C ₂)

(6) 统计分析(具体分析过程参见第 4 章 4.5 节)。通过 SAS 程序，得出如表 3-20 所示的试验结果方差分析表。

表 3-20 试验结果方差分析表

方 差 来 源	df	SS	MS	F 值	P 值
A	1	0.0666	0.0666	1.67	>0.05
B	1	0.4851	0.4851	12.19	>0.05
C	1	0.0021	0.0021	0.05	>0.05
AB	1	0.0561	0.0561	1.41	>0.05
AC	1	0.0595	0.0595	1.49	>0.05
误差	2	0.0795	0.0398		
合计	7	0.7490			

由表 3-20 可以看出，以 0.05 的检验水准，各个因素的 P 值均大于 0.05，即各个因素均为显示出对大白鼠的细胞色素 b5 有影响。

值得一提的是：上面这个例子实际上是无重复试验的 3 个 2 水平因素的一种安排方法，相当于未做重复试验的“三因素析因设计结构”。其结果的正确性取决于以下两点：其一，BC 和 ABC 这两个交互作用项在客观上是可以忽略不计的；其二，各次试验（即正交表上每一行）条件下的试验结果的波动在专业上允许范围之内，否则，必须进行重复试验。

9. 序贯试验设计

医学科研设计一般都是预先确定样本含量，结果分析在全部试验完成后做出。序贯试验（sequential trial）预先不需要确定试验例数。事先根据试验标准绘好序贯试验图。将每次试验结果的数据点到图上，当达到要求的检验界限时，便可停止试验，做出判断。

【例 3-17】 用猫试验氯丙嗪的抗呕吐效果，给每只家猫用氯丙嗪后再用吐根碱，观察是否呕吐，并且规定抗呕吐率大于 60% 为有效，小于 30% 为无效， $\alpha = \beta = 0.05$ ，观察分析氯丙嗪是否有抗呕吐效果。

【实例辨析】 $\pi_1 = 60\%$ $\pi_0 = 30\%$ 为衡量抗呕吐有效无效的标准，且呕吐与否能及时观察到，故适宜用序贯设计。由于本例只问氯丙嗪是否有抗呕吐作用，故为单向试验，每只猫的结果只有呕吐或不呕吐两种，故属质反应。记录的试验结果如表 3-21 所示。

表 3-21 氯丙嗪对猫抗呕吐试验结果

编 号	1	2	3	4	5	6	7	8
呕吐反应	-	+	-	+	+	+	+	+

分析过程如下。

(1) 规定试验标准。

有效率 $\pi \geq \pi_1 = 60\%$ 时，处理有效。

有效率 $\pi < \pi_0 = 30\%$ 时，处理无效。

$\alpha = \beta = 0.05$

(2) 计算处理因素有效及无效两界限的直线方程。

上界 $U: y = a + bn$

下界 $L: y = -a + bn$

$$a = \frac{\log \frac{1 - \beta}{\alpha}}{\log \frac{\pi_1 (1 - \pi_0)}{\pi_0 (1 - \pi_1)}} = \frac{\log \frac{1 - 0.05}{0.05}}{\log \frac{0.6 (1 - 0.3)}{0.3 (1 - 0.6)}} = 2.35, \quad b = \frac{\log \frac{1 - \pi_0}{1 - \pi_1}}{\log \frac{\pi_1 (1 - \pi_0)}{\pi_0 (1 - \pi_1)}} = 0.45$$

上界 $U: y = 2.35 + 0.45n$; 下界 $L: y = -2.35 + 0.45n$

(3)将两界限的直线方程绘制成图，横轴代表样本量，纵轴代表有效样本。

(4)绘制试验线：第一例试验若有效，向正东北方向画一格，无效，向正东画一格，第二例试验从第一例所画线段的终点起，按相同规定画线，如此序贯进行。最终得到的图形如图 3-1 所示。

总之,动物试验费时较长，消耗财力和物资较大，操作烦琐、比较复杂，因此既要遵循严谨的科研精神，又不宜墨守成规，要有一定的灵活性，切实做好预测，周密设计剂量，精心饲养动物，细心观察反应，认真测定指标，实事求是总结。

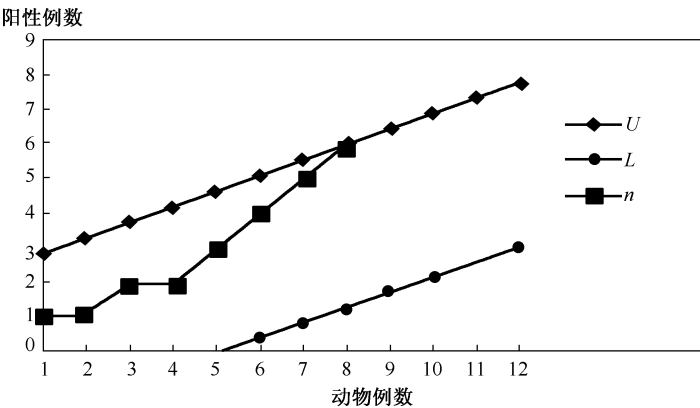


图 3-1 序贯试验设计图

3.4 动物试验常见问题及其对策

在内科学的发展过程中，动物试验是不可或缺的一种研究途径。人们通过解剖试验来了解动物体本身的形态和结构，通过受控的动物试验来了解各种外界因子对动物生理及行为的影响，并从而为将这些因子作用于人和其他动物而做好前期的工作。随着动物保护运动的广泛开展及伦理学视野的扩大，动物试验的正当性和必要性、试验动物的保护等问题逐渐凸现出来，成为学术界和社会关注的重点话题。

3.4.1 我国动物试验领域存在的问题

1. 饲养和试验环境落后

我国许多大学一级动物试验室建设已有较高水平和规模。但是，动物试验室的发展还比较滞后，房舍陈旧，设施更显落后，有的七八年竟未添置过仪器设备。试验动物不达标，饲养员也未经专业训练。大多是雇来的临时工，给试验动物喂多喂少、营养搭配随意性很大，根本谈不上科学性。

研究表明动物的生活环境会影响到其大脑发育。在老鼠身上做试验时,科学家发现，生活在空间较大且附加了迷宫、梯子、海绵等刺激物的笼子里的老鼠，其大脑中乙酰胆碱的含量相对较高，而通过这种酶可以看出神经系统的活动情况。

试验证明，不同的居住环境会改变神经细胞体的大小和突触接触的数量。而使用那些神经系统有问题的动物做试验，无论是涉及行为科学、神经科学还是神经解剖学，研究者都可能被引入歧途，影响到研究所得出的数据的有效性和准确性。这意味着疾病模型的设计、药剂研究和化学毒性检测的结果都将受到影响。

2. 动物试验的操作情况不佳,有虐待动物的现象

在一份对医学生的调查中,试验前对试验操作步骤“非常熟悉”的人仅占10%，“熟悉”的占52.99%，而“不是很熟悉、需要一边看步骤一边试验”的达到36.51%。

动物试验大部分是活体试验,学生在试验前对试验步骤及注意事项不清楚,会延长试验时间,延长动物的痛苦,甚至导致试验失败,动物也就白白做了牺牲。这说明部分学生缺乏严谨的科学态度。这也表现在动物的麻醉给药上,约10%的学生对动物进行麻醉给药时是“只凭估计给药”或“随便给药”。这种不科学的给药方式是对动物痛苦的无视和对动物生命的淡漠。缺少严谨的科学态度,一切试验结果就有可能失真,在动物试验上就成了对动物的虐待。

3.4.2 我国动物试验替代方法(3R)研究存在问题

所谓3R理论,即在动物试验中遵循减少(reduction)、替代(replacement)和优化(refinement)的原则。“减少”是要求在试验中尽可能减少试验动物的使用数量,提高试验动物的利用率和试验的准确性;“优化(精细)”是确保动物在麻醉、镇痛、镇静或其他适当的手段作用下进行试验,不使其遭受不必要的伤害或痛苦。“替代”是不再利用活体动物进行试验而是以组织细胞培养、各种活体外试验或计算机模型以及统计分析等方法来加以替代。

(1)从科研管理角度讲,没有建立起统一的规划指导。

我国目前还没有专门的法规,在应用动物进行试验的各个不同领域缺乏有关规定和法律,可以说,试验动物替代方法在我国的试验动物和有关动物试验的法规建设中还处在比较薄弱的位置上,与较发达国家相比,差距较大。

(2)研究经费明显不足。

我国开展“3R”研究的科研经费主要来自政府管理部门,因科研发展水平以及其他各方面原因,“3R”研究还不可能作为科研体系中的主项,没有稳定的经费渠道支持相应研究工作,经费支持力度也无法与其他领域的研究计划和科研项目相比较。

3.4.3 针对我国动物试验问题的对策

(1)注意在动物试验过程中动物伦理的实施。为了尽量减少动物在试验过程中的痛苦和惊恐,我们应严格按照操作规程的要求,对动物轻拿轻放,试验中注意对动物进行抚慰,待动物安静后方进行备皮、染毒等操作。进行动物保定时,尽量使动物保持相对舒适的姿势,一旦试验结束,迅速解除保定,从而最大限度地减少了动物的不适。

(2)改善试验动物的饲养环境,对试验动物房进行了重新设计和建设,解决以往普通级动物房排风不畅,易留死角,温、湿度难以控制的问题。减少了动物房内的异味,大大改善了动物的生存环境,保障试验动物在饲养过程中能够得到相应的福利待遇。

(3)加强动物试验伦理教育,培养敬畏生命的意识。医药院校虽普遍开展了生命伦理教育系列活动,但很少有专门针对动物试验的主题活动。在伦理学课堂上,教师很少系统地讲授动物试验的伦理问题,再加上在操作试验前,任课教师也很少主动强调,从而导致了学生在动物试验中得到的相关人道主义教育严重缺失。试验室是医学生成长的第一步。医学专业教师与伦理学老师需要共同努力,从细小处着手才能达到医德培养的最佳效果。为此,医学院校应重视3R理论基础上的动物试验伦理教育,在伦理学的教学方面可增加动物试验伦理专题,让学生从理论上深入了解与思考;在组织生命伦理主题活动时,可组织类似珍爱试验动物的活动,让学生从实践中体会敬畏生命的医学人文内涵。

(4)努力实践“3R”,完善相关法律法规。在西方一些国家,动物试验替代方法的研究已达到较高的水平,专门的研究机构、多种渠道的资金来源、不同层次的学术刊物都为“3R”研究和应用的

进一步深入和广泛开展打下了良好的基础。

据此,我们应该积极学习西方一些国家,建立专门的基金会和研究机构,用于“3R”理论的普及、教育和“3R”技术的研究、验证和应用。以“3R”所提倡的指导性原则作为国家立法和科研工作的基础。

生命科学的研究者们对待动物要发扬人道主义精神。在科学研究上,求真是非常重要的一个目标,但求善也是绝对不可以被忽略的。在客观的基础上激发科学家的人文情感,才能在寻求真理的道路上走得更好。

附例 1 内科学动物试验设计主要包括以下基本步骤

1. 开始阶段

(1)文献查阅。查阅的文献应广泛包括现在和过去的杂志以及网络上的信息资源,如 MEDLINE、TOXLINE、PUBMED、NCBI 等,目的是了解清楚你对所研究的焦点问题都已经知道些什么,明确所用的方法、动物模型,排除不必要的重复研究。也是为 IACUC 审查提供项目替代动物试验的方法是否可行、可能引起动物疼痛的操作已得到评估认可的证据。

(2)科研方法。观察和描述试验中的现象,系统地提出问题的解释和学说,利用学说预测新的观察结果,验证学说。

(3)陈述问题预期可能的结果。明确提出问题十分关键。提出的问题应包括试验将要说明的问题和它的意义。对每一个试验方案有两个清晰的结果,如一个无效,另一个备择假说。无效假说明确两试验组间没有差别,备择假说表明在试验组之间有真实的差别。尽管可能随着试验的进行对这些问题的可能加以修改,但明确提出问题和假说是进行试验设计的基础。

例如:

- ① 提出问题:哪种饲料引起大鼠体重增加,是饲料 A 还是 B?
- ② 无效假说:各组试验预期得到同样结果,即大鼠用饲料 A 和 B 体重增加相同。
- ③ 备择假说:试验组之间预期出现不同的结果。
- ④ 不可检验的假说:结果不容易明确或解释。
- (4)动物模型的确定。为试验选择最适当的动物模型,有以下几点建议:
 - ① 使用系统发育水平最低的动物。
 - ② 使用的动物具有想要的种属或品系专一的特点或者具有特定研究目的必需的特点。
 - ③ 考虑在试验期间得到和维持动物模型所需要的花费。
 - ④ 充分查阅文献,与同一领域的同事讨论,与供应商或动物模型资源库联系确定动物模型可能的来源渠道;
 - ⑤ 在最终决定动物模型前征询试验动物兽医的意见。

(5)试验人员的确定:根据试验实施的操作,确定需要具有什么样专长的人。在开始项目前,就确定每个人的职责和进行的工作以及资格。合理地进行人员的分工安排,参与计算、试验设计和样品收集对确保所得数据的有效性是十分重要的。一般一个大型研究单位的试验动物中心设施可提供涉及本专业高技术含量的操作和昂贵装备的服务。

2. 正式设计阶段

(1)研究方案。针对提出的问题、研究目标和假说,进行试验的所有操作及安排细心地计划并记录在案。方案中应包括针对可控试验变量的试验操作、适合测定的变量参数(即得到准确测定试验变量控制的效应)、获得样品和产生试验数据的最适方法。项目的所有实物和数据收集和测定的时间安排表都应在这时确定。

实际操作的讨论包括动物模型持续的时间(慢性研究)、模型中预期疾病的进程(决定测定的最适时间点)、人员参加项目的时间和岗位、试验花费。已知和潜在的危害因子也必须明确,最小危害因子风险适当预防措施也必须在方案中明确。所有试验操作步骤应通过 SOP 详细规定。方案中还应包括数据分析的方法,如果组间差异是通过统计学计算而得,那么应明确所用统计学检验的方法。由数据分析对最初所提问题做出回答并接受或否决假说,由此得出结论。

(2) 试验单元。研究的实体是试验单元,可以是一个动物个体,也可以是一组动物群体。例如:当测试药物治疗和外科手术时,一个大鼠个体就是一个试验单元;当测试环境致畸时,一整窝大鼠就是一个试验单元。

(3) N 因子。试验组别的动物只数,将适当数量的动物分配到每一个试验组别,是试验设计的一个重要环节。为了得到统计学意义的结果,每一试验组的动物数量一般是通过预期各组间结果差异的可能程度采用特定统计学检验而决定的。一个标准笼所能容纳动物的数量也是试验组动物数量确定的实际考虑的因素。

(4) 对照。过多的变量(如遗传、环境、感染因子)对动物试验的结果产生影响,因此设置对照动物消除这些外来变量或可能未知的变量带来的影响。总之,每一个试验都应使用对照且对照组动物与试验组动物有一个直接对应的关系。对照的种类很多,包括阳性对照、阴性对照、空白对照、媒介物对照、比拟物对照。

(5) 随机化。动物分组时必须将动物随机分配到试验各组,以确保研究的变量在各组中不会导致偏差数据。

为达到随机化,必须在开始时将动物群体限定好。随机化的常用方法:

- ① 每一只动物有一个固定的编号,进行“暗箱”抽数,并随机分配它们到不同组中。
- ② 利用随机数字表或计算机随机程序。

3. 设计后阶段

(1) 试验操作步骤的确认。IACUC 对试验操作的确认,其他委员会对试验操作的确认。

(2) 人员的确认。参加研究项目的所有人员必须经过培训和资格认可。

(3) 小规模预试验。是使用数量少的动物得到预示性的数据,目的如下:

- ① 通过预试验将操作和使用的技术固定完善。
- ② 预试验的数据是申请资助项目方案中证明其合理性的重要依据,增加中标的机会。
- ③ 估计变量参数测定的误差,为计算样本量提供依据。

(4) 数据的输入和分析。研究者对数据的正确收集、输入和分析负有完全的责任。当处理的数据量很大时,很容易发生数据输入的错误(如组标识调换、个体标识调换)。因此,在分析数据前,应建立数据质量保证程序并在试验设计中包含,确保数据输入的错误能被及时辨别。

(5) 评价。

附例2 诱发性肿瘤动物试验设计

1. 试验设计名称

诱发性肿瘤动物试验设计。

2. 试验方法原理

诱发性肿瘤动物试验设计是指研究者用化学致癌剂、放射线、致癌病毒诱发动物的肿瘤等。

3. 试验设计的目的

- (1) 评价抗肿瘤免疫治疗的疗效。
- (2) 作为抗肿瘤药物筛选模型。
- (3) 为肿瘤转移研究提供更好的研究平台。
- (4) 为研发抗肿瘤转移性药物提供良好的试验工具。

4. 试验所用材料

试验所用材料包括肿瘤细胞、体重 250g 左右的封闭群大白鼠(雌雄均可)、无血清培养基质 3% 中性甲醇石蜡、低温离心机血球计数器游标卡尺等。

5. 试验过程

(1) 肿瘤诱发。

- ① 取体重为 250g 左右的封闭群大白鼠(雌雄均可) 10 ~ 20 只。
- ② 设计不同的试验组和对照组进行试验对照, 每组动物数一般为 5 ~ 10 只, 施行配对设计。
- ③ 按性别分笼饲养。两组全部给予相同的普通食物、饲养环境保持一致。试验组另饲以致癌物, 即用 0. 25% 二乙基亚硝胺(DEN)水溶液灌胃, 剂量为 10mg/kg, 每周一次, 其余 6 天用 0. 025% DEN 水溶液放入水瓶中, 任其自由饮用。
- ④ 连续喂养 4 个月或直至试验组诱发出肝癌。

(2) 数据采集及处理。

- ① 一般接种肿瘤 6 ~ 8 周后, 小鼠的肿瘤可生长至直径为 15 ~ 20mm(即小鼠会濒临死亡)时, 可眼球取血, 分离血清并保存血清, 处死小鼠, 取肿瘤组织照相, 取部分肿瘤组织作为冰冻组织切片(或 - 80℃ 保存), 做相应的免疫组织化学染色, 取部分肿瘤组织用 3% 中性的甲醛固定, 石蜡包埋, 做常规 HE 染色。
- ② 记录肿瘤大小、生长时间、甲胎蛋白、谷氨酰转肽酶等指标数据。

③ 对数据进行 χ^2 检验。

(3) 注意事项。

① 亚硝胺类致癌性强, 小剂量一次给药即可致癌; 对多种动物(包括猴、豚鼠等不易诱发肿瘤的动物)的许多器官(包括食管、脑、鼻窦等不易引起癌的器官)能致癌, 甚至可以通过胎盘致癌, 如给怀孕大鼠以二乙基亚硝胺可比较快地引起仔鼠的神经胶质细胞瘤; 具有不同结构的亚硝胺有明显的器官亲和性, 例如二甲基亚硝胺等对称的衍生物常引起肝癌, 不对称的亚硝胺如甲基苄基亚硝胺常诱发食管癌; 在大鼠, 二丁基亚硝胺能引起膀胱癌, 二戊基亚硝胺能诱发肺癌, 而 N-甲基-N-硝基-N-1-亚硝基胍则能引起胃肠癌。因此在进行操作时应注意剂量控制, 防止出现其他副反应从而而影响试验效果。

② 以前的动物试验已确定: 乙硫氨酸可诱发大鼠肝癌; 乌拉坦(氨基甲酸乙酯, $C_2H_5O-CONH_2$)、氨乙胺可引起小鼠的肺癌; 一些卤代烃如四氯化碳、氯仿等可引起大鼠或小鼠的肝癌。因此在进行试验设计时, 要注意防止其他因素混入而引起试验效果偏倚。

附例 3 著名医学科学家的重大发现与动物试验的关系

(1) 哈维(William Harvery, 1578—1657) 是英国医生, 试验生理学的创始人之一。他采用狗、蛙、蛇、鱼、蟹和其他动物进行了一系列动物试验。根据大量的试验研究结果, 发现了血液循环, 证实了动物体内的血液循环现象, 并阐明了心脏在此过程中的作用, 指出血液受心脏推动, 沿动脉

流向全身各部,再沿静脉返回心脏,环流不息,他还测定过心脏每搏输出量。他在1628年发表《动物心血运动的解剖研究》,1651年发表《论动物的生殖》,这些成就对生理学和胚胎学的发展起了很大作用。恩格斯对哈维的发现给予了高度的评价:“由于哈维发现血液循环,而把生理学确定为一门科学。”

(2)科赫(Robert Koch,1843—1910)是德国细菌学家。他采用牛、羊和其他动物做试验,发现了结核杆菌。他发明用固体培养基的“细菌纯培养法”,首先采用染色体观察细菌的形态,并运用这些方法,分离出炭疽杆菌、结核杆菌和霍乱杆菌,同时确证这些细菌与疾病的关系,提出了“科赫原则”,作为判断某种微生物是否为某种疾病的病原的准则,于1905年获诺贝尔生理或医学奖。

(3)巴斯德(Louis Pasteur,1822—1895)是法国微生物学家、化学家、近代微生物学的奠基人。他在病原微生物方面的研究,奠定了医学微生物学的基础。在研究蚕病,鸡霍乱和炭疽病中,证实传染病是由病原微生物所引起的。他采用鸟类做动物试验,发现被减毒的鸡霍乱和炭疽病原菌能诱发免疫性。他晚年在鸟和家兔上进行狂犬病疫苗的研究,对狂犬病免疫做出了很大贡献。他在研究炭疽病中,证实传染病是由病原微生物所引起的,其中有一个生动的事例。巴斯德很想知道有的地方为什么不断发生炭疽病,而且总是发生在同样的田野里,有时相隔数年之久。巴斯德从埋了12年之久、死于炭疽病的羊尸体周围土壤中,分离出这种病菌。他奇怪这种有机体为什么能这样长时间地抗拒日照以及其他不利因素。一天,巴斯德在地里散步时,发现一块土壤与周围颜色不同,遂请教农民。农民告诉他说,前一年这里埋了几只死于炭疽病的羊。一向细心观察事物的巴斯德注意到土壤表层有大量蚯蚓带出的土粒。于是,他想到蚯蚓来回不断从土壤深处爬到表层,就把羊尸体周围富有腐殖质的泥土以及泥土中含有的炭疽病芽孢带到表层。巴斯德从不止步于设想,他立刻进行了试验,试验结果证实了他的预见,接种了蚯蚓所带泥土的豚鼠得了炭疽病。

(4)巴甫洛夫(1848—1936)是俄国生理学家,他一生做了大量的动物试验,在心脏生理、消化生理和高级神经活动三个方面做出了重大贡献。他在早年发现温血动物心脏有特殊的营养性神经,能使心脏增强或减弱。在消化腺的研究中,他在狗身上创造了许多外科手术,改进了试验方法,以慢性试验代替了急性试验,从而能够长期地观察整体动物的正常生理过程,在研究消化生理过程中,形成了条件反射的概念,从而开辟了高级神经活动生理学研究,他的高级神经活动学说对于医学、心理学以及哲学等方面都有很大影响。1904年,他获诺贝尔生理或医学奖,著有《动物高级神经活动(行为)客观研究二十年经验》等著作。由于他在研究中经常不断地使用狗作研究对象,因此他的一些著作也以狗来命名,如《狗的血压的正常变动范围》、《狗的神经的支配》等。他对动物试验给予了高度的评价,如“没有对活动物进行试验和观察,人们就无法认识有机界的各种规律,这是无可争辩的。”他对试验动物的作用和习性也很了解,有很多精辟的论述,如“狗由于素来对人好感,由于它的机敏、耐性以及驯顺而十分愉快地为试验者服务许多年,甚至终身。”“只有当必须时才用猫做试验,因为这种动物性情急躁,本性凶恶,善叫。”“除了狗以外,家兔是最常用的试验动物,因为它是一种驯顺而活泼的动物,而且很少尖叫与反抗。”

(5)贝尔纳(Claude Bernard,1813—1878)是法国生理学家,他的最重要的发现为肝脏的产糖功能和血管运动神经。他观察到刺激第四脑室底部能使动物发生暂时性的糖尿症,表明体内糖的产生是受中枢神经系统控制的。胰液能分解中性脂肪的功能和美洲箭毒能麻痹骨骼肌肉的作用机制也是由他发现的。他区别有机体的外环境和内环境,他所指的内环境主要是血液。他认为尽管外环境不断变化,内环境却保持恒定是生命的保证,这个概念启发了后来生理学的许多的研究。贝尔纳在上述一些重大发现中,做了大量动物试验,其中有几个典型例子,如贝尔纳要根据冲动沿交

感神经传导并引起化学变化从而在皮肤中生热的假说,切断了家兔颈部的交感神经,希望导致兔耳变凉。使他吃惊的是:该侧的耳朵却变得更热了。贝尔纳将耳血管与通常使耳血管保持适当收缩的神经作用彼此脱离了联系,结果血液流量增大,耳朵变热。贝尔纳起初并没有认识到自己的所作所为,他完全偶然地发现了动脉中的血流量是由神经控制的这一事实,这是自哈维经典性的发现以后,人类对血液循环认识的最重要进展之一。又如,有一天,别人给贝尔纳的试验室送来了几只从市场上买来的兔子。贝尔纳注意到试验桌上兔子排的尿清亮而带酸性,不像寻常草食动物那样混浊而带碱性。他推断,这多半是由于没有喂食,兔子从自己身体的组织中吸取养分,因而处于食肉动物的营养状况。他用喂养和禁食互相交替的方法证实了这个观点,这种作用过程果然使兔尿反应发生了预期的变化。这是一次精彩的观察,多数研究人员也就心满意足了,但贝尔纳却不然。他要求“反证”,于是用肉食喂兔子,果然不出所料,兔尿呈酸性,贝尔纳为完成这项试验,最后对兔子做了解剖。用他自己的话说“我偶然注意到白色乳状的淋巴液初见于离幽门约 30cm 处十二指肠下部的小肠中。这引起了我的注意,因为在狗的身上淋巴液初见于十二指肠的上部紧靠近幽门的。”再仔细观察,他看到胰导管的开口是与淋巴液开始含有白色乳糜的位置一致,脂肪质的乳状液使这种乳糜成为白色,这样就发现了胰液在脂肪消化中的作用。

(6) 莱夫勒(Friedrich Löffler, 1852—1915) 是德国细菌学家,在白喉研究的早期,证明了试验动物因注射白喉杆菌而死亡时,细菌仍留在注射点的附近。他认为动物死亡是由细菌的毒素所造成的。根据这一假说,鲁(Emile Roux, 1853—1933, 法国细菌学家、医生)做了大量动物试验,企图证实细菌培养液中的这种毒素,虽做了很多努力,却都失败了。尽管如此,鲁仍坚信这一假说,最后孤注一掷,给豚鼠注射了 35mL 的大剂量培养液滤液,奇怪的是,这只豚鼠在注射了如此大剂量的液体居然没有立即死亡,过了一些时候,他满意地看到这只豚鼠死于白喉中毒。确认了这点以后,鲁很快就查明,这只豚鼠开始之所以不死是因为培养液中细菌培养时间不够长,产生的毒素不足所致。因此,增加细菌培养时间就能够制成毒性很大的滤液,这一发现导致了预防白喉的免疫法,并使抗血清用于治疗(1890 年, Von Behring 发现患白喉病愈后豚鼠的血清中含有某种物质,可以保护其他豚鼠),从而开始了抗毒素治疗的新的时代。

(7) 冯梅林(Baron Joseph Von Mering, 1849—1908) 是德国内科医生;闵可夫斯基(Oskar Minkowsk, 1858—1931) 是出生于俄国的内科医学、病理学家。1889 年,他们在斯特拉斯堡研究胰脏在消化过程中的功能时,用手术切除了一个狗的胰脏。过后,一个试验助手发现这只狗的尿招来了成群的苍蝇。他将此事报告给了闵可夫斯基,经分析尿后发现其中有糖。正是这一发现,使我们认识了糖尿病和后来用胰岛素控制糖尿病的方法。胰岛素是由 Banting 和 Best 从家犬体内首次分离出的,这一发现拯救了无数糖尿病患者的生命。

(8) 里基特(Charles Ricet) 是法国生理学家。他在用试验动物试验海葵触手提取物以测定其毒素剂量时,突然发现,与第一次相隔一段时间,第二次的微小剂量常使动物迅速死亡。起先,他对此大为震惊,简直不能相信这是他自己做出来的结果。确实,他说过,他发现诱导敏感作用或称过敏性完全是不知不觉的,他原来认为这是绝对不可能的。这种过敏现象的另一表现是由戴尔(Henry Dale)发现的。他在豚鼠的几条不随意肌肉内注射血清时,突然发现有一条肌肉对马血清反应特别强烈。在寻找这一特别现象的原因时,他发现这只豚鼠在不久前曾注射过马血清。由于前次已注射过马血清(称感应剂量)使豚鼠对马血清(异性蛋白)的感受提高,因此第二次再注射马血清(称决定性剂量)时,豚鼠可比第一次注射反应强烈得多。这证实了过敏的本质是抗原抗体的反应,从而推动了变态反应性疾病的研究。

(9) 洛伊(Otto Loewi) 是格茨茨大学药理学教授。1921 年,他以创造性思维,仅采用简单的动物试验方法,就发现了副交感神经的神经介质为乙酰胆碱。次日,他走进试验室,以生物学历史上

少有的利落、简单、肯定的试验证明了神经冲动的化学媒介作用。他准备了两只蛙心，用盐水使其保持跳动。他刺激一只蛙心的迷走神经，使其停止跳动。然后他把浸泡过这只蛙心的盐水取出来浸泡第二只蛙心。洛伊满意地看到，盐水对第二只蛙心的作用，同刺激迷走神经对第一只蛙心的作用相同：搏动的肌肉停止了跳动。这就是世界各国对化学媒介作用不仅存在于神经与它们影响肌肉和腺体之间，而且也存在于神经单元本身之间。

(10) Ringer's 液的发现：林格(Sidney Ringer, 1835—1910)是英国内科医生、生理学家。在对离体的青蛙心脏进行试验时，生理学家通常使用生理盐水作为灌液。用这种方法可使青蛙心脏继续保持约半小时的跳动。一次，在伦敦大学医院，一位生理学家发现他的青蛙心脏连续跳动了好几个小时，他非常惊讶，大惑不解，他能想到的唯一可能原因是季节的影响，而对这一点，他也确实在报告中提出来了。后来，发现这是由于他的试验助手在制作盐水溶液时用的不是蒸馏水而是自来水。根据这个线索断定自来水中的某些盐分引起了生理活动的增加。林格就是这样发现了这种以他名字命名的溶液。这种溶液对试验生理学的贡献颇大。

(11) 脾结节法测定多向造血干细胞方法的发现：1961年，J. E. Till 和 E. A. McCulloch 两位加拿大血液学家在研究正常小鼠骨髓细胞放射敏感性时，对受致死剂量射线照射的小鼠移植一定数量的同种骨髓细胞后 9~10 天，取出脾脏，在苦味酸-甲醛固定液中固定后，可以肉眼见到脾脏表面上生成圆形结节。从而创立了脾结节法。同时，他们进一步观察到脾结节的生成量与移植骨髓的有核细胞数之间呈比例关系，因此，可以作为评定骨髓细胞功能的一个重要测定方法，为测定造血细胞的功能提供了一个定量的研究方法。后来，进一步的工作证明，脾结节的生成是起源于单一细胞的，这类细胞具有很强的增殖能力，同时又具有向骨髓红系、粒系和巨核系细胞进行分化的能力，符合了造血干细胞的基本特征，因而称脾结节的生成细胞为多向性造血干细胞或多潜能性造血干细胞。因此，脾结节法是目前公认的多向性造血干细胞的测定方法，它的发现大大促进了血液学的研究。

(12) 其他一些著名科学家采用动物试验获得重大发现的例子还很多，如化学致癌物质的发现：1914年，日本人山极和市川用沥青长期涂抹兔耳朵，成功地诱发出皮肤癌，后经分析沥青中主要含有 3, 4-苯并芘的化学致癌物，从而证实了化学物质的致癌作用。从此，许多化学物质都相继被证实可经诱发动物的肿瘤，为肿瘤病因的化学因素提供了更多的证据。使人们充分认识到化学致癌因素在人类恶性肿瘤的病因中占有极重要的地位。现在已知的有致癌作用的化学物质的种类繁多，分布广泛。例如 3, 4-苯并芘，1, 2, 5, 6-双苯蒽，20-甲基胆蒽等多环碳氢化合物，均被动物试验证明是强烈的致癌物质。用这些物质涂抹动物皮肤可引起皮肤癌，注射在动物皮下则可引起肉癌。氨基偶氮染料，可引起动物肝癌和膀胱癌。芳香胺类染料，可引起膀胱癌，其中芳香酰类(如 N-3-乙酰氨基苄)可使多种动物引起肝、肠、乳膜、外耳道、膀胱等不同部位的肿瘤。亚硝胺类化合物是一类分布广，致癌性强，能引起多种动物及多种器官发生肿瘤的天然存在的致癌物，甲基苄基亚硝胺、甲基丙烯基亚硝胺等不对称的亚硝胺主要引起食管癌；二甲基亚硝胺、二乙基亚硝胺等对称亚硝胺，主要引起肝癌；甲基亚硝基脲和甲基亚硝基乌拉坦能诱发大鼠的小肠、肝、肾、脑的肿瘤和胃的腺癌。

(13) 单克隆抗体技术的发明：杂交瘤(Hybridoma)合成单克隆抗体(Monoclonal antibodies)是近年来生物医学中的一项重大突破，从根本上解决了免疫学中长期存在的“特异性”和“重复性”问题，显出它的发展前途无可限量。单克隆抗体技术是由英国剑桥大学两位科学家 G. Kohler 和 C. Milstein 于 1975 年发明的，他们在 20 世纪 60 年代发展起来的细胞杂交技术基础上，成功地把两种细胞融合在一起：一种是已适应体外培养的小鼠骨髓瘤细胞(都来自 BALB/C 品系的小鼠)和一种经绵羊红细胞免疫的小鼠脾细胞融合，形成杂交细胞，发现这种融合的杂交细胞兼有两

个亲代细胞的特征，即既有骨髓细胞无限生长的能力，又有浆细胞合成单一抗体的能力。因此，这种免疫细胞通过克隆化，成为单克隆系(单一纯化的无性繁殖系)，就能产生大量单一类型的高纯度抗体，这种抗体就称为“单克隆抗体”。如果把杂交细胞再种入动脉体内，可形成“杂交细胞瘤”瘤体产生大量抗体，就可从动物体液中抽出含有单克隆抗体的体液，把这种杂交细胞在体外培养，培养液中就有大量的单克隆抗体，可供试验研究和临床研究使用。最近，在探索用于融合的人体浆细胞瘤方面已获得成功，为单克隆抗体直接用于人类疾病的诊断、预防、治疗以及发病机制的研究，特别是为人类恶性肿瘤的免疫诊断、免疫治疗，开辟了更为广阔的前景。现在，单克隆抗体技术几乎深入到生物医学的各个领域，甚至渗入到过去认为免疫血清学方法达不到的范畴，具有无法估量的作用。

(安洪庆 王素珍)

第 4 章 内科临床试验设计与数据分析

4.1 内科临床试验的基础知识

4.1.1 临床试验的定义

临床试验 (clinical trial) 是以病人或健康志愿者为研究对象, 比较干预 (intervention) 与对照 (control) 所显示的效果及其临床价值的一种前瞻性研究。常用来评价药物、疗法、技术和预防性干预措施的效果。

评价药物的临床试验, 通常用于证实或揭示试验药物的作用、不良反应或试验药物的吸收、分布、代谢和排泄, 以确定药物的疗效和安全性。评价不同疗法的疗效是将某种新的治疗方法在部分病人中试用, 与一般方法或无处理对照组病人进行比较, 根据治愈率、降低伤残率和病死率等指标对其临床疗效做出科学的估计。评价预防性干预措施的效果时, 在实施预防性干预组和无处理对照组之间比较, 根据发病率、发病者症状轻重等方面来判断。

与动物研究不同, 在临床试验中研究者不能完全支配病人的行为, 只能要求病人避免某些干扰试验的行为。临床试验中除应遵循试验设计的基本原则和方法外, 在病例选择、设立对照、疗效指标的选择、避免试验偏差及效果评价方面均有其特点。

4.1.2 临床治疗试验的特点

1. 以人为研究对象

临床试验的研究对象可以是病人 (干预措施评价), 也可以是健康人 (预防、预后效果评价)。人具有社会属性, 受精神因素、心理因素影响, 与动物试验比较, 外来影响因素更加难以控制。试验必须在保证病人安全的前提下进行, 应符合医学伦理要求。

2. 试验过程受多个因素影响

由于受试对象是既具有生物学属性, 又具有社会属性的病人或志愿者, 在研究过程中, 研究者不可能完全支配他们的行为, 更不可能控制他们的思维和精神心理活动, 试验观测条件也很难完全加以控制。为了避免不同处理组病人的心理差异, 或其他治疗措施以外的干扰因素 (如疾病的自然缓解、自愈倾向等) 对疗效判定的影响, 采用随机化分组、设立对照组的方法, 保证非试验因素分布均衡。

3. 干预措施的实际效果要全面评估

临床试验的效果可以表现为多个方面, 并且相互联系。评估干预措施的实际效果时, 既要注意近期效果, 又要追踪机体康复状态等远期效果。根据全面、完整的资料进行综合评价, 得出正确结论。

4. 受到伦理道德的规范

受试对象在得到必要和足够的信息并充分理解了这些信息后, 经过对这些信息的考虑, 自由地

做出了参加研究的决定,而没有受到任何强迫、威胁、诱导或不正当影响,即参加临床试验的受试对象是知情同意的,知晓试验中自己的受益之处且具有退出的自由。试验项目要得到伦理委员会的批准。

4.1.3 临床试验的分期

《国家药品注册管理办法》规定,临床试验分为Ⅰ、Ⅱ、Ⅲ、Ⅳ期,新药在批量生产、投放市场前,除了进行动物药理试验、毒理试验等基础试验外,还必须按规定进行临床Ⅰ、Ⅱ、Ⅲ期临床试验。

1. Ⅰ期临床试验 (phase Ⅰ clinical trial)

Ⅰ期临床试验(phase Ⅰ clinical trial), 又称临床前期试验或小规模临床试验, 包括人体耐受性试验和药物代谢动力学试验。这一期试验提供了人体对于新药的耐受程度和药代动力学的的数据, 为制定该药方案提供依据。

人体耐受性试验: 根据动物试验的有效量、耐受量和中毒量, 估算出符合安全要求的最小剂量及最大剂量, 为Ⅱ期临床研究推荐合适的剂量。通常从预定的有效量的 1/10 或从动物试验最小有效量的 1/30 ~ 1/20 开始, 逐步增加剂量到体内出现明显效应为止。

人体药物代谢动力学试验: 根据人体耐受性试验确定的三个剂量, 在人体(主要是健康志愿者)测定血中药物浓度的历时性变化, 进行药物代谢动力学分析, 计算有关参数。了解新药的吸收、分布、消除的特点, 为Ⅱ期临床研究提供合适的给药方案。该期试验人数一般为 20 ~ 30 例。由于其数量少, 无对比, 研究又“公开”, 所得结论具有不确定性。

2. Ⅱ期临床试验 (phase Ⅱ clinical trial)

Ⅱ期临床试验(phase Ⅱ clinical trial), 又称正式临床试验。主要评价药物疗效, 其目的是初步评价药物对适应症患者的疗效和安全性评价, 也为Ⅲ期临床试验的给药剂量方案提供依据。通常, 用随机对照试验设计方案进行设计。对给药的适应症、给药方法、药理和毒理反应及药物的真实效果等进行深入对比研究。评价联合用药的效应时, 可考虑采用析因设计。该期试验组人数不少于 100 例。

3. Ⅲ期临床试验 (phase Ⅲ clinical trial)

Ⅲ期临床试验(phase Ⅲ clinical trial), 又称多中心临床试验。在 3 个或 3 个以上指定的医院同时进行随机对照试验, 即多中心临床试验。其目的是进一步验证药物对适应症患者的有效性和安全性, 评价利益与风险关系, 最终为药物注册申请获得批准提供充分的数据。这是治疗效果的确证阶段; 是扩大的对比临床观察, 广泛地观察新药的各种适应症、不良反应、毒性和药理作用。该期试验组人数一般不少于 300 例。

4. Ⅳ期临床试验 (phase Ⅳ clinical trial)

Ⅳ期临床试验(phase Ⅳ clinical trial) 是新药上市后由申请人自主进行的应用研究阶段, 即新药上市后对药物疗效和安全性的进一步监测。样本数量不限, 被观察地区、人群广泛, 采取观察性研究方法, 观察期限可长达数年或更长。其目的是考察在广泛使用条件下的药物的疗效和安全性, 评价在普通或者特殊人群中使用的利益和风险关系, 改进给药剂量等。该期试验组人数一般不少于 2000 例。

4.1.4 临床试验的统计分析集

在临床试验的统计分析中, 哪些受试者应当包括在内, 哪些受试者不应当包括在内, 这是分析试验结果时首先要考虑的问题, 即“分析集”(analysis set) 问题。根据纳入分析的受试对象的不同, 分析集分为四类。

1. 意向性分析(intention to treat, ITT)原则

意向性分析(intention to treat, ITT)是指主要分析所有随机化的受试者,按其所分到的组别进行随访、评价和分析而不管其是否依从计划完成过程。此原则保证了原始的随机化分组,为统计分析提供可靠的基础。但实际工作中有时是不合理的:比如,随机化分组后试验组受试者吃了对照组的药物或在随机化后没有任何资料。

2. 全分析集(full analysis set, FAS)

全分析集(full analysis set, FAS)是指尽可能按意向性分析原则,由所有随机化的受试者以合理的方法尽可能少地、合理地剔除受试者所形成的数据集。

3. 符合方案集(per protocol set, PPS)

符合方案集(per protocol set, PPS)是指全分析集中更加符合方案的受试者子集,一般包括完成最小剂量的药物治疗、依从性好、完成主要疗效指标测量的受试者所形成的数据集。

4. 安全性评价集(safety set, SS)

安全性评价集(safety set, SS)是指用于安全性评价的数据集,应包括所有随机化后至少接受一次治疗的受试者所形成的数据集。

4.2 内科临床试验设计的要素及原则

4.2.1 内科学试验设计的三要素

临床试验属于试验性研究范畴。因此,试验设计的基本要素和试验研究相同,即试验因素、受试对象和试验效应三个组成部分。对于临床试验来说,试验因素就是所要评价、比较的临床干预措施(诊断措施、治疗措施或预防措施)。受试对象就是接受这些干预措施的病人或志愿者。试验效应就是在受试对象身上观察到的临床效应。例如,用两种降压药治疗高血压病人,观察比较服用不同降压药的两组病人血压值的下降情况,这里所用的降压药就是试验因素,高血压患者为受试对象,被测的血压值则为试验效应。构成医学试验的这三个基本组成部分称为医学试验的三个基本要素。任何一项试验研究在进行设计时,必须首先明确这三个要素,并以其为重点制定详细、严密的研究计划。

临床试验是以人为受试对象的,在试验设计中还需要对以下两方面问题给予充分考虑:

- (1)必须考虑医学伦理道德,保证试验不会给受试对象带来伤害。
- (2)必须考虑到受试对象和研究者本人精神心理因素对观测结果带来的影响。

1. 干预措施

临床试验的特点之一是研究者能人为地设置干预措施。为确保正确、恰当地设置干预措施,一般应注意以下几点。

- (1)临床干预措施的合理确定。

每个临床试验都要至少确定一种要观察评价其临床效应的干预措施。这种干预措施首先应是对受试对象有利无害的,还必须是可以通过现有观测手段观测到其临床效应的。对上述两点一般要求有临床前药理学、毒理学的动物试验依据。只观察和评价一种干预措施临床效应的称为单因素临床试验研究,同时观察和评价两种和两种以上干预措施临床效应以及它们之间交互作用的称为多因素临床试验研究。虽然在研究的深度和广度上,多因素临床试验研究都优于单因素临床试验研

究,但在一次临床试验中不宜安排过多的干预措施,否则不仅会增加试验设计和数据分析处理的难度,还会分散研究精力,甚至造成主次不分、喧宾夺主的情况。

干预措施的临床效应往往与干预措施的强度或水平有关,即存在着“量—效”关系。因此,在确定干预措施的同时还要把施加的强度或水平合理地确定下来。干预措施的强度不够可能产生不出明显的临床效应;强度过高又可能出现较大的副作用或毒性反应,危及受试者的安全,必须结合专业知识和临床前研究结果积极稳妥地加以确定。

(2) 找出可能影响试验结果的非试验因素。

非试验因素虽然不是研究者关注的试验因素,但其中有些可能会影响试验结果。当它们在试验因素不同水平组中所产生的影响不相等时,就产生了混杂效应,此时的非试验因素又被称为混杂因素。例如上面提到的用两种降压药治疗高血压病人的试验,非试验因素可能有年龄、性别等,不同的年龄、性别,降压效果可能不一样。若两种降压药组的年龄、性别构成不同,则可能影响降压药疗效的正确评价。设计时明确了这些非试验因素,才能设法消除或降低它们的干扰作用。通常,设法使试验组与对照组中的患者年龄的平均值和标准差分别相同或相近;男女性别构成相同。

(3) 干预措施必须标准化。

干预在整个试验过程中应做到标准化,即保持不变,否则会影响试验结果的评价。例如,一项对某新药疗效评定的临床试验,试验的干预措施是新药物,那么该药物的成分、剂型、规格、生产厂家、批号、所用剂量、给药方式、使用程序等必须保持不变,并要在临床试验设计方案中做出明确的规定。

2. 受试对象

临床试验的受试对象(study subjects)不同于一般的动物试验,因为受试对象是特殊的群体——“人”,即病人或健康志愿者,因此其选择尤为重要,对试验结果有着极为重要的影响。

(1) 符合试验的目的要求。

所选的病人或志愿者必须是对所施加的干预措施敏感者,能够显现出可观测到的临床效应。作为病人,诊断必须明确无误,符合公认的、统一的诊断标准(如果没有公认的统一诊断标准,则要根据专业理论自行制定出统一的诊断标准,而且必须有客观指标作为诊断依据)。作为志愿者必须是真正的志愿者,而不是由于某种特定原因,如经济困难、病情危重、无其他有效治疗措施而被迫成为“志愿者”的,以免带入选择性偏倚。

(2) 有较高的同质性。

符合统一的诊断标准是选择受试者的首要条件,但不是唯一条件。每项临床试验都有其特定的研究目的,需要特定的研究对象。因此,还要根据具体的研究目的对符合统一诊断标准的对象做进一步“纯化”,制定出受试对象入选标准和排除标准。把不符合试验要求或不适合接受预定干预措施的人排除出去,把真正符合试验目的要求的对象接纳进来,使入选的受试者具有更好的同质性,从而更有效地排除要观察的试验因素之外的其他非试验因素对试验观测结果的干扰,得出更加可信的结论;同时也可以尽可能地减少不良事件的发生。例如,一项考核某抗菌药物疗效的临床试验规定的受试对象入选标准是:①患有急性细菌性感染需要进行全身抗菌药物治疗的病人;②年龄在18~65岁之间;③细菌学检查检出了对该抗菌素敏感的菌株。排除标准是:①患有严重肝、肾、心脏及造血系统疾病;②有严重过敏史;③妊娠及哺乳期妇女;④有明显精神症状无法合作者;⑤正在应用其他抗菌素药物者。一般来说,受试对象的“纯化”程度越高、同质性越好、试验的齐同可比性就越高,所得出的试验观测结论说服力也就越强。也应看到,受试对象“纯化”程度越高,其代表表面就越窄,研究结果的推广应用范围就越局限,同时选择受试对象的难度也越大。所以入选和排除标准既不可定得过宽,也不宜定得过严,应根据研究目的和实际情况妥善加以制定。

国家卫生部颁发的中、西新药临床研究指导原则中,对部分新药临床病人纳入标准和排除标准进行了规定,研究者可参照执行。

受试对象可分为病人或健康人。病例的选择应当诊断明确,依从性好,能真实、可靠地反映主观感觉。

(3)有良好的依从性。

依从性(compliance)是指受试者在试验过程中对于试验因素的服从程度,即受试对象忠实执行医嘱的程度。例如,能否按时接受治疗、是否按规定服药、能否自始至终与研究者的密切合作、能否真实地反映和表述临床效应、由于各种原因退出试验的可能性,等等。依从性差常常是导致不能正确评估临床效应的重要原因。因此,临床试验应该尽量选择依从性好的病人作为受试对象。一般来说,住院病人的依从性比门诊病人的依从性好,也更便于监督管理。如果以门诊病人作为受试对象,则要注意避免选择那些可能中途退出试验的病人。在临床试验设计方案中要制定出保证高依从性的具体措施,如尽可能简化治疗方案,缩短疗程,定期随访,送医送药上门,以优良的服务态度、优质的服务水平促使受试对象依从。对负担医疗费用有困难的受试对象可酌情减免费用等。

(4)签署知情同意书。

每个受试对象都了解试验的目的要求、预期效果和潜在风险,自主决定是否参加试验。这也是提高依从性的一项重要措施。对自愿参加试验、符合入选条件的对象要请其签署知情同意书。同意书中必须申明受试者有随时退出试验的权利。

3. 临床试验效应

临床试验效应(clinical effect)是受试对象对干预措施的反应,它通过观察指标来表达。试验效应是研究结果的最终体现,也是试验研究的核心内容。效应主要通过客观指标和主观指标来表达,因此,要想使试验因素的作用得到准确反映,必须选择恰当的指标。指标分为客观指标和主观指标两大类。客观指标包括定量指标、定性指标及等级指标,主观指标是受试者的感觉或研究者的主观判断。同时,在选择指标时,也应考虑指标的客观性、灵敏性、特异性和精确性。

(1)指标的选择。

试验效应的观察指标应当选择客观性强、灵敏度高和精确性好的指标。

①客观性。客观指标是测量和检验的结果,比主观指标更为可靠。当指标有客观指标和主观指标时,应尽量选择客观指标,少用及慎用主观指标,如在评价某药对胃溃疡的疗效时,用胃镜检查结果作为观察指标比患者对疗效的评价作为观察指标更加客观。但有的主观指标也越来越得到重视和应用,如生命质量指标就在大量研究中得到广泛的应用。

②灵敏性和特异性。指标的选择应具有较高的检出真阳性的能力(灵敏度高),例如在研究某药物治疗缺铁性贫血的效果时,血清铁蛋白就是一个灵敏度高的指标;待选指标也要有较高的鉴别真阴性的能力(特异度高),例如甲胎蛋白对于原发性肝癌就是特异性较高的指标。因此,在选择指标时,应同时考虑指标的特异性和灵敏性问题。

③精确性。指标的精确性包括准确度和精密度两层意思。准确度(accuracy)指观察值与真值的接近程度,主要受系统误差的影响。精密度(precision)指重复观察时,观察值与其均数的接近程度,其差值属随机误差。试验效应指标要求既准确又精密,两者皆优的指标,若两者存在矛盾时,则优先考虑准确度。

(2)指标的观察。

临床效应的观察应避免带有偏性(bias)。例如研究者的心理常偏重于阳性结果,医生常偏重于新疗法组,病人常对新疗法持怀疑态度等。这些都可能导致试验效应指标测定时带有偏性,从而影响结果的比较和分析。为了消除或减少这种测量偏差,在设计时常采用盲法(blind method)。如果

受试对象不知道设计者的方案，哪组是试验组，哪组是对照组，预期结果是什么，称为单盲法。若受试对象及试验执行者均不知道设计方案，称为双盲法，在临床试验中用得比较广泛。

4.2.2 内科学试验设计的基本原则

同理，临床试验属于试验研究的范畴，除了符合试验研究的三项基本原则外，还要符合盲法的原则。另外，由于以人为研究对象，故应考虑病人或志愿者的知情同意、心理因素、伦理道德等问题。

1. 对照原则

对照原则指在确立接受试验因素的试验组时，同时设立不接受试验因素的对照组(control)，从而显出试验因素的作用。试验组与对照组除了试验因素不同外，其他条件应相同或接近，即均衡性好。这样才能平衡非试验因素对试验结果的影响，把试验因素的效应充分显示出来。临床上有很多疾病和症状，如感冒、慢性气管炎、腰背扭伤、关节酸痛和早期高血压等不经药物治疗，也会自愈或随着季节变化而缓解。建立对照组的目的是为了消除其他因素的作用，因此，对照是控制混杂因素的基本措施。对照的方式有多种，可依研究目的和内容选择，常用的有以下几种。

(1)空白对照。

空白对照(blank control)是一种不施加任何处理的对照。其在动物试验以及试验室方法研究中常采用空白对照。例如，在某种可疑致癌物的动物诱癌试验中，设立与试验组动物种属、窝别、性别和体重相同或相近动物作为空白对照。以排除动物本身可能自发肿瘤的影响。空白对照有时也用于人群试验。例如观察阿司匹林预防心肌梗死的临床研究，试验组口服阿司匹林，对照组不服，追踪观察一定时间后，比较2组心肌梗死的发生率，即属于空白对照组。空白对照简单易行，但容易引起试验组与对照组在心理上的差异，从而影响试验效应的测定。只适合于受试对象为未患病的志愿者和病情相对稳定或不予治疗也可自愈的病人，如感冒、慢性气管炎、近视眼患者等。临床试验中，在涉及伦理道德问题时，不宜提倡空白对照，要慎用。

【例4-1】 空白对照实例或错误辨析：观察黄连素对试验性Ⅱ型糖尿病小鼠脂代谢异常的影响水平。方法：制造Ⅱ型糖尿病小鼠模型20只，分成治疗组与空白对照组，同时另取10只设立正常对照组。治疗组每天用黄连素溶液(100mg/kg)灌胃1次，空白对照组及正常对照组在相同时间灌服相同容积生理盐水，疗程30d，治疗结束取血测血脂及血胰岛素。结果：治疗组血TC、TG、LDL-C较空白对照组明显降低($P<0.01$)，HDL-C较空白对照组明显升高($P<0.05$)，如表4-1所示。结论：黄连素能明显改善试验性Ⅱ型糖尿病小鼠的脂代谢异常。

表 4-1 三组小鼠血脂、血胰岛素指标结果比较 ($\bar{x} \pm s$)

组 别	<i>n</i>	TC mmol/L	TG mmol/L	HDL-C mmol/L	LDL-C mmol/L	血胰岛素 μU/mL
治疗组	10	4.16 ± 0.77	2.07 ± 0.40	1.85 ± 0.37	1.84 ± 0.81	46.29 ± 7.20
空白对照组	8	6.72 ± 2.54	3.87 ± 1.40	0.64 ± 0.28	4.42 ± 1.13	58.52 ± 3.83
正常对照组	10	3.49 ± 0.23	1.99 ± 0.26	1.31 ± 0.22	2.76 ± 1.02	42.38 ± 4.66

【实例辨析】 本题的主要目的是观察黄连素对试验性Ⅱ型糖尿病小鼠脂代谢异常的影响水平，在设置的3个组中，空白对照组是多余的，其实此处的正常对照组实际为试验对照，故本题只需要有治疗组和试验两组即可。额外设置空白对照，不仅造成人力、物力、财力资源的浪费，而且也容易造成比较不同组时，尺度把握得不一致，因此，此处应去掉空白对照组。

(2) 安慰剂对照。

安慰剂对照 (placebo control) 是对照组采用一种无药理作用的假药, 其在药物剂型和处置上不能为受试者识别, 称为安慰剂。口服的安慰剂常用乳糖、葡萄糖、维生素、淀粉等成分加工制成。注射用安慰剂常用生理盐水。使用安慰剂有助于避免对照组病人产生与试验组病人不同的心理作用, 但在使用时一定要慎重, 做好“保密工作”。一般来说, 适用于小规模的研究。

【例 4-2】 安慰剂对照实例或错误辨析: 研究中医药治疗对冠心病介入术后病人生命质量的影响。方法: 将 59 例进行了介入治疗的冠心病病人, 随机分为试验组 (30 例) 和对照组 (29 例), 试验组以西药常规治疗加用中药通冠胶囊, 对照组以西药常规治疗加用安慰剂; 应用 SF-36 生命质量量表对病人介入术前后进行评价, 观察不同组别之间的改善情况。结果: 治疗 1 月后, 试验组 SF-36 量表评分中身体疼痛、精神健康、生理功能、总体健康等方面及总分均优于对照组 ($P<0.01$ 或 $P<0.05$), 如表 4-2 所示。结论: 中医药在冠心病介入术后病人中应用可以提高病人的生命质量。

表 4-2 两组治疗前后 SF-36 计分比较 ($\bar{x} \pm s$)

组	别	n	生 理 功 能	生 理 职 能	身 体 疼 痛	总 体 健 康
试验组	治疗前	30	67.09 ± 13.01	24.98 ± 11.69	60.85 ± 22.69	62.38 ± 18.86
	治疗后	30	65.11 ± 19.01	26.32 ± 15.09	58.79 ± 25.46	60.89 ± 10.97
对照组	治疗前	29	79.05 ± 11.06	31.19 ± 10.63	89.38 ± 19.63	82.01 ± 11.63
	治疗后	29	76.96 ± 12.39	29.96 ± 19.07	69.06 ± 11.68	79.86 ± 9.18

组	别	活 力	社 会 功 能	情 感	精 神 健 康	健 康 变 化
试验组	治疗前	54.77 ± 11.72	43.26 ± 19.32	51.25 ± 16.34	44.00 ± 30.68	55.09 ± 11.36
	治疗后	53.86 ± 12.12	44.09 ± 20.13	50.29 ± 20.19	41.96 ± 27.36	53.86 ± 11.00
对照组	治疗前	69.36 ± 10.18	65.29 ± 11.28	90.15 ± 10.29	81.06 ± 16.09	78.08 ± 5.49
	治疗后	67.95 ± 8.16	63.72 ± 10.35	85.06 ± 6.00	78.06 ± 19.06	63.09 ± 8.01

【实例辨析】 本研究的目的是观察中医药治对冠心病介入术后病人生命质量的影响。研究者将介入治疗的冠心病病人随机分成试验组和对照组, 对照组在西药的基础上加安慰剂, 试验组在西药的基础上加中药通冠胶囊, 最后两组比较反映出中药胶囊是否改善冠心病介入术生命质量。对照组安慰剂的加入, 消除病人心理上的顾虑。从而使两组的试验效应变量测的数据更均衡可比。

(3) 试验对照。

试验对照 (experimental control) 是对照组不施加试验组中的真实处理, 但施加某种与试验组有关的“本底”。例如一项观察超声疗法对急性化脓性乳腺炎疗效的临床试验。试验组在服用抗生素的同时给予超声治疗。对照组只服用抗生素, 以保证组间的均衡性。在此例中, 给予相同种类和剂量的抗生素是“本底”, 而试验因素是“是否给予超声治疗 (其两个水平分别为是与否)”。

【例 4-3】 试验对照实例或错误辨析: 探讨麝香注射液治疗老年脑梗死的可能机制。方法: 以红细胞免疫酵母菌花环试验方法为基础, 将复方麝香注射液药物作为干预因素, 根据加入或不加入复方麝香注射液分为药物干预的试验观察组、空白对照组和试验对照组, 分别对其红细胞免疫指标 RBC-C3bRR、RBC-ICR 进行试验对照分析。结果: 如表 4-3 所示, 空白对照组、试验对照组的 RBC-C3bRR 与试验观察组 RBC-C3bRR 相比较, 均有统计学意义 ($P<0.05$)。而空白对照组与试验对照组相比较无统计学意义 ($P>0.05$)。3 组 RBC-ICR 相比较均无统计学意义 ($P>0.05$)。结论: 麝香注射液对脑梗死病人红细胞免疫功能有明显的正向调节作用。

表 4-3 不同试验方法对 RBC-C3bRR 及 RBC-ICR 的调控作用 ($\bar{x} \pm s$) %

组 别	n	RBC-C3bRR	RBC-ICR
试验观察组	50	14. 61 \pm 2. 78	14. 30 \pm 3. 35
试验对照组	50	8. 00 \pm 1. 49	14. 80 \pm 3. 11
空白对照组	50	8. 16 \pm 1. 42	14. 30 \pm 2. 88

【实例辨析】 本研究中使用脑梗死患者的新鲜肝素抗凝血，制备血浆。在此基础上加入复方麝香注射液，等容积的生理盐水，什么都不加入，完成试验观察组、试验对照组、和空白对照组三组样品的制备，在此基础上进行红细胞免疫酵母菌花环试验，在显微镜下计数 100 个红细胞，以黏附 2 个或以上酵母菌为一阳性花环，分别计算百分率为 RBC-C3bRR、RBC-ICR。最后比较各组的 RBC-C3bRR、RBC-ICR 以达到研究目的。此处的试验对照组没有施加试验因素，却加入了和处理无关的试验因素，即等容积的生理盐水，故为试验对照。而一管脑梗死患者血浆样品中什么都没有施加，故为空白对照。

(4) 标准对照。

标准对照 (standard control) 是用公认的标准疗法或常规疗法作为对照。多用于评价一种干预措施与公认有效的处理措施的临床效应有无差别。例如，新药临床试验，试验组给予新药，对照组给予公认有确切疗效的传统药物。标准对照试验根据研究目的又可分为优效性试验 (superior trial) 和等效性或非劣效性试验 (equivalence trail or non-inferiority trial)。前者目的是评价试验组的临床效果是否优于标准对照组。后者目的是评价试验组的临床效果是否与标准对照组相当 (既不优于也不劣于) 或不比标准对照组差。

(5) 潜在对照。

潜在对照 (potential control)，又称历史对照或文献对照，指不专设对照组，而是以公认的已发生的事实为对照。例如某癌症过去无一例治愈，若现研制出某种新药，临床医生用该药治愈了几例该癌症患者，在此药物试验中，尽管未专门设立对照组，而用潜在的以前该癌症无数个不可治愈的例子作为对照。故恶性肿瘤、狂犬病、艾滋病等不能治愈是公认的事实，可采用潜在对照的形式。除了难以治疗的疾病外，一般不提倡潜在对照。因为随着时间的推移，医学科学技术的进步，以前不可治愈的疾病变得可治愈，历史资料和现存资料可能存在不可比性问题。

【例 4-4】 潜在对照实例辨析：1963 年 1 月 2 日，上海市第六人民医院陈中伟医师等为一名 27 岁的青年工人成功地接活了一只完全断离的右手前臂，半年后，再植肢体功能已恢复良好功能。这是世界上第一次断肢再植获得成功。

【实例辨析】 由于世界上从未有过断肢再植成功的案例，因此，仅此 1 例子成功的例子也能说明我国已经具备先进的显微外科医疗技术。

【例 4-5】 2006 年 4 月 13 号，第四军医大学西京医院整形外科主任郭树忠教授等开展了世界第二例，国内第一例“换脸手术”，经过长达 17 个小时的异体颜面移植术，成功解决了血管吻合与血液供应，骨骼、软骨、肌肉、腮腺等组织对接，出血与感染控制等一系列手术关键技术。

【实例辨析】 此手术取得了巨大成功，与法国开展的世界首例异体颜面移植手术比较，其包含的组织量更多。因此，“换脸手术”也说明我国异体颜面移植术处于国际领先水平。

【例 4-6】 世界首例全面换脸手术获得成功，一名为奥斯卡的男子于 2010 年 3 月在巴塞罗那的瓦尔德希伯伦大学附属医院接受了手术。五年前，奥斯卡发生射击事故，包括鼻子和嘴唇在内的原有面部表皮和肌肉大部分遭受损害，呼吸、吞咽、说话等最基本面部功能都难以发挥，因此不得不寻求所谓换脸手术。最终，在巴莱特大夫 (J. P. Barret) 的率领下，一组医生经过长达 24 小时的一系列神经肌肉和美容手术，成功地向患者面部重植肌肉、鼻子、嘴唇、上颌骨、软口盖、全部牙

齿、脸颊骨等面部功能器官。这是世界上首例全面换脸手术。

【实例辨析】 由于世界上从未有过全面换脸手术成功的案例，因此，仅此 1 例子成功的例子以潜在的对照为基础说明了该项技术已经成功。

(6) 自身对照。

自身对照 (self control)，即试验与对照在同一批受试对象中进行。每个受试对象既是试验对象，也是对照对象，这样既节省样本，又易控制条件，还可以有效地消除个体差异的影响，提高试验效率。在具体实行中又有以下几种形式。

①自身平行对照。在同一受试对象的不同部位分别给予试验处理和对照处理，平行观察临床效应，如用左侧肢体做试验，右侧肢体做对照；左眼做试验，右眼做对照等。

②自身前后对照。用同一病人接受处理前的情况作为接受处理后情况的对照。这种对照的设置简单易行，容易为患者接受，又不存在医学伦理学问题，故临床上应用颇广。但严格地讲，它不是随机分配的，有一定的局限性。不宜用自身前后对照的情形：对短期自愈倾向或有周期性发作倾向的疾病；试验前后某些环境因素或自身因素发生了改变，并且会影响试验结果时；时间间隔过长，观察指标自身可能会发生明显改变。

【例 4-7】 自身前后实例或错误辨析：观察心脑血管舒通片治疗椎-基底动脉供血不足性眩晕 (VBA) 的临床疗效。方法：54 例 VBA 供血不足性眩晕患者被随机分成治疗组和对照组。治疗组口服心脑血管舒通片每次 2~3 粒，每日 3 次，对照组口服西比灵 5~10mg，每晚 1 次。疗程为 2 周。对比观察其临床疗效及经颅多普勒超声 (TCD) 变化。结果：两组治疗前后收缩压情况如表 4-4 所示，两组治疗前后收缩期血流速度的变化如表 4-5 所示。可见，治疗组和对照组都能有效地改善患者的临床症状和 TCD 结果，两组在临床疗效上的差异无统计学意义。结论：心脑血管舒通片治疗 VBA 供血不足性眩晕安全有效。

表 4-4 两组治疗前后舒张末期血流速度的变化 ($\bar{x} \pm s$) cm/s

组 别		VA(L)	VA(R)	BA
治疗组	治疗前	13.40 \pm 1.14	13.06 \pm 1.64	16.02 \pm 1.39
	治疗后	21.52 \pm 1.44	18.76 \pm 1.56	25.20 \pm 1.90
对照组	治疗前	13.18 \pm 1.26	13.11 \pm 1.51	13.77 \pm 1.29
	治疗后	17.41 \pm 1.14	17.06 \pm 3.30	18.20 \pm 0.92

注：与本组治疗前比较，* 代表 $P < 0.01$ 。

表 4-5 两组治疗前后收缩期血流速度的变化 ($\bar{x} \pm s$) cm/s

组 别		VA(L)	VA(R)	BA
治疗组	治疗前	30.40 \pm 3.04	30.66 \pm 2.57	37.94 \pm 2.79
	治疗后	39.92 \pm 2.73	38.76 \pm 2.86	47.20 \pm 2.90
对照组	治疗前	29.78 \pm 2.45	34.12 \pm 3.47	42.73 \pm 3.22
	治疗后	36.40 \pm 2.14	35.06 \pm 3.33	45.20 \pm 3.16

【实例辨析】 本研究中给治疗者服用心脑血管舒通片，对照组服用西比灵，然后将治疗组治疗前后表明舒张末期血流速的指标和收缩期血流速指标相比较，对照组也将这两类指标相比较。无论治疗者和对照组都是同一受试对象服用药物前后得到两次数据的比较，因此，两组均为自身对照。如果研究目的仅为观察心脑血管舒通片治疗椎-基底动脉供血不足性眩晕 (VBA) 的临床疗效，仅对治疗组服用心脑血管舒通片前后的舒张末期血流速的指标和收缩期血流速指标进行比较即可。

③成组交叉对照。是一种特殊的自身对照，即将受试对象随机分为 2 组。给予第一组试验处

理，给予第二组对照处理，间隔一段时间(称为洗脱期)后再改为给予第一组对照处理，给予第二组试验处理，这样使每个受试对象既接受过试验处理又接受过对照处理。最后把接受试验处理时产生的临床效应与接受对照处理时产生的临床效应进行对比，得出结论。这样既节省样本含量，又易控制试验观测条件。但要求每次处理后产生的临床效应能在短时间内消失，受试者会重新“恢复”到处理前状态。故多用于慢性病或神经官能性疾病的疗效评定。

【例 4-8】 成组交叉对照实例或错误辨析：观察 40mg/kg 番泻叶浸液预防治疗化疗后便秘的疗效。方法：将 82 例化疗后曾出现过便秘的患者分为 AB 与 BA 组，AB 组指第 1 个化疗周期为服用番泻叶浸液，第 2 个化疗周期为服用粗纤维饮食组；BA 组则相反，观察番泻叶浸液及粗纤维饮食对化疗后缓解便秘的有效率，结果如表 4-6 所示。结果：番泻叶浸液对便秘总有效率为 92.68%，粗纤维饮食为 10.93%，两者比较差异有统计学意义($P < 0.01$)。结论：40mg/kg 番泻叶浸液预防治疗恶性肿瘤化疗后便秘，具有较好疗效。

表 4-6 82 例患者两种疗法比较

组 别	例 数					例 数		
	* : 0	I	II	III	IV	# : I	II	III
番泻叶浸液	60	14	2	0	0	3	3	0
粗纤维饮食组	8	1	0	0	0	12	54	7

注：* 代表“腹泻分级”，#代表“便秘分级”。

【实例辨析】 本研究中采用成组交叉对照设计，所有患者按入组先后顺序编号，奇数号进入 AB 组，偶数号进入 BA 组，成组交叉对照，共入组 82 例患者。对 82 例患者均观察 2 个化疗周期(化疗方案及所用止吐药相同)，AB 组第 1 个化疗周期服用番泻叶浸液，第 2 个化疗周期服用粗纤维饮食；BA 组则相反，化疗第 1 个周期和第 2 个周期之间间隔 2~3 周，作为洗脱期。采用成组交叉设计节约了样本含量，且两组也更均衡可比。值得一提的是，本例在整理资料时，并没有按交叉设计的要求去做，而是按单因素两水平设计整理的。

(7)相互对照。

相互对照(reciprocal control)，指受试对象随机分为几个组，每组分别接受一种干预措施或者每组接受同一种干预措施的不同水平，观察比较其临床效应有无差别。如果已知几种干预措施都是有效的，则可不再另设对照，各组间互为对照。如果给予不同水平的同一种干预措施观察比较其临床效应，多用于药物临床试验，确定最优剂量。剂量组中可以包括零剂量(zero-does)，也可以不包括零剂量。

【例 4-9】 相互对照实例或错误辨析：观察慢性肾功能衰竭(CRF)维持性血液透析(HD)患者血浆总同型半胱氨酸(tHcy)的影响因素和大剂量叶酸治疗是否可以纠正其高同型半胱氨酸血症(HHcy)。方法：77 例稳定的 HD 患者被分为 4 组，分别口服叶酸 0、5、15 或 30mg/d，3 个月以上(所有患者均不使用维生素 B₁₂)。用放射免疫分析法测定透析前和血浆叶酸、全血叶酸和维生素 B₁₂。用高效液相法测定透析前和部分患者透析后血浆同型半胱氨酸(tHcy)。用常规方法测定总胆固醇(TC)、甘油三脂(TG)、高密度脂蛋白(HDL)、低密度脂蛋白(LDL)、血清白蛋白(Alb)、肌酐(Ser)、尿素(Urea)。观察不同剂量的叶酸治疗对血浆和全血叶酸的影响，以及血浆和全血叶酸与 tHcy 的关系。结果：如表 4-7 所示，可以看出，叶酸治疗可以明显提高血浆和全血叶酸水平。口服叶酸 5~15mg/d 均可使血浆叶酸大于 200ng/ml，此时血浆和全血叶酸的进一步上升不引起 tHcy 的进一步下降。如下表所示。结论：对 HD 患者，治疗 HHcy 的叶酸的合适剂量是 5~15mg/d，进一步增大叶酸剂量并不能使 tHcy 水平进一步下降。口服叶酸结合 HD 治疗不能完全纠正 HD 患者 HHcy。

表 4-7 叶酸治疗对 sFA、tFA、VitB₁₂、tHcy 的影响

组 别	例 数	sFA (ng/mL)	tFA (ng/mL)	VitB ₁₂ (pg/mL)	tHcy (μmol/L)
对照组	19	5.21 ± 0.84	94.84 ± 9.17	431.65 ± 120.94	6.36 ± 5.43
非治疗组	15	11.4 ± 13.0	735.1 ± 282.2	1346.8 ± 1712.5	41.4 ± 25.2
5mg/d 组	15	221.2 ± 258.3	1737.6 ± 867.5	1192.2 ± 1669.8	28.9 ± 7.24
15mg/d 组	17	661.9 ± 318.1	3306.0 ± 877.8	1193.5 ± 1835.1	24.6 ± 8.67
30mg/d 组	30	560.3 ± 452.7	3686.9 ± 1600.1	1235.6 ± 1431.0	25.4 ± 6.99
正常参考值	—	3 ~ 17	80 ~ 320	200 ~ 950	< 15

【实例辨析】 本研究中将 77 例随机分为 4 组，分别口服叶酸 0、5、15、30mg/d 达 3 个月以上，同时测定同期 19 例正常健康人(无进食叶酸或维生素 B₁₂ 史)空腹 sFA、tFA、VitB₁₂、tHcy 作为对照组。观察不同剂量叶酸对 sFA、tFA 及 tHcy 的影响。各组互为对照进行比较，采用方差分析的方法比较各组之间 sFA、tFA 及 tHcy 之间是否存在统计学差异。结果表明不同组(15mg/d 组与 30mg/d 组可视同为大剂量组)之间 sFA、tFA 和 tHcy 差异有统计学意义， $P < 0.05$)。严格地说，本例中“组别”之下的 5 个组不是单因素五水平设计，若对任何一个定量指标进行差异性分析，最好将“组别”拆开，形成两种不同组合，再判断其试验设计类型。例如，对照组与非治疗组比较，属于成组设计；后四组一起比较，属于单因素四水平设计问题。

2. 随机化原则

随机化(randomization)是用随机的方式抽取样本，使符合条件的试验对象参加试验的机会相同，同时将受试单位随机地分配到各处理组中，使每个受试单位被分到试验组与对照组的机会均等。随机化使非试验因素在试验组与对照组中的影响较为一致，是保证比较组间有良好均衡性的重要手段之一。可见，随机不等于随便，既不是由受试者随意选择，也不是由研究者主观决定，而是通过随机化的分组程序来实现的。常用的随机化分组方法如下。

(1) 完全随机分组。

该分组方法是利用随机数对受试对象分组，且不要求各组例数相等。其基本步骤如下。

①给受试对象编号。将 n 个受试对象按一定特征顺序从 1 到 n 依次编号，患者可按顺序就诊。

②分配随机数字。随机数字可从计算机或计算器上产生，或由随机数字表读取。从随机数字表读取时，可由表中任一个数字开始，沿一个方向顺序给每个受试对象抄取一个随机数字，随机数字的位数一般要求与 n 的位数相同。

③按随机数字所在区间分配相应受试对象所在的组别。例如，按 2 位随机数字分两组时，可规定随机数 00 ~ 49 为第一组，50 ~ 99 为第二组；若分 3 组，可取 01 ~ 33 为第一组，34 ~ 66 为第二组，67 ~ 99 为第三组，余类推。分两组时，也可由随机数的奇、偶决定组别。

上述分组方法不能保证各组别例数相等，若要求各组例数相等，可将随机数字按由小到大的顺序排列序号，再按排列序号 R 进行分组。

【例 4-10】 试用完全随机分组方法，将 18 个孕妇分配到 3 组，分别服用 A、B、C 3 种保健药物，观察新生儿脐血胰岛素水平有无差异。

①18 个孕妇按就诊顺序依次编序号 1 ~ 18。

②从随机数字表中任意选择起始数，现从第 10 行第 11 列向右读取随机数，依次读取 18 个 2 位随机数，对应抄在编号下，见表 4-8。

③规定随机将 01 ~ 33 对应的孕妇分到甲组，34 ~ 66 分到乙组，67 ~ 99 分到丙组，结果见表 4-9。

表 4-8 18 个孕妇完全随机分组方案

编 号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
随机数	18	46	23	34	27	85	13	99	24	44	49	09	79	74	16	32	02	57
组别	甲	乙	甲	乙	甲	丙	甲	丙	甲	乙	乙	甲	丙	丙	甲	甲	甲	乙

由表4-8 可见，甲组9 例，为1、3、5、7、9、12、15、16、17 号孕妇；乙组5 例，为2、4、10、11、18 号孕妇；丙组4 例，为6、8、13、14 号孕妇。3 组例数不同。要求3 组例数相等时，可将上述随机数从小到大排序，得序号 R ，并规定 R 是1~6 的为甲组， R 是7~12 的为乙组， R 是13~18 的为丙组，结果见表4-9。

表 4-9 18 个孕妇完全随机分组方案

编 号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
随机数	18	46	23	34	27	85	13	99	24	44	49	09	79	74	16	32	02	57
R	5	12	6	10	8	17	3	18	7	11	13	2	16	15	4	9	1	14
组别	甲	乙	甲	乙	乙	丙	甲	丙	乙	乙	丙	甲	丙	丙	甲	乙	甲	丙

由表4-9 可见，甲组孕妇编号为：1、3、7、12、15、17；乙组孕妇编号为：2、4、5、9、10、16；丙组孕妇编号为：6、8、11、13、14、18。

如果采用 SAS 程序进行单纯随机化，程序如下：

```
proc plan seed = 20130922;
  factors a=18;
  output out = sheji;
run;
data b1 b2 b3;
  set sheji;
  patient = _n_;
  if a < =6 then do; group = 'A'; output b1;end;
  if 6 < a < =12 then do; group = 'B'; output b2;end;
  if 12 < a < =18 then do; group = 'C'; output b3;end;
run;
data d1;
  set b1;
  group1 =group;
  patient1 =patient;
  keep group1 patient1;
run;
data d2;
  set b2;
  group2 =group;
  patient2 =patient;
  keep group2 patient2;
run;
data d3;
  set b3;
  group3 =group;
  patient3 =patient;
  keep group3 patient3;
run;
```



```
data e;
merge d1 d2 d3;
run;
proc print noobs;
run;
```

程序说明：第一步，调用 PLAN 过程产生 18 个随机排列数，并把结果输出至数据集 sheji 中，选项 seed = ，规定产生随机数的种子。第二步，对数据集 sheji 进行操作，把患者按随机排列数进行分组，随机数小于等于 6 的进入 A 组，大于 6 小于等于 12 的进入 B 组，大于 12 小于等于 18 的进入 C 组，并将结果分别保存至数据集 b1、b2、b3 中。第三、四、五步，分别将数据集 b1 b2 b3 中的 group 改为 group1、group2、group3，将 patient 改为 patient1、patient2、patient3，生成新的数据集 d1 d2 d3。最后把数据集 d1 d2 d3 合并为新的数据集 e，并通过 PRINT 过程将随机分组的结果展现出来。

SAS 运行结果如下：

group1	patient1	group2	patient2	group3	patient3
A	1	B	7	C	3
A	2	B	10	C	5
A	4	B	11	C	6
A	9	B	13	C	8
A	15	B	14	C	12
A	18	B	16	C	17

(2) 区组随机排列方法。

区组 (block) 由若干特征相似的试验单位组成。如病情严重程度相同，年龄、体重相近等受试者都可组成一个区组。区组随机排列是将每一区组内的处理顺序随机排列。其方法是先将各区组内的观察单位按某特征排序，同时也将处理规定顺序，如 A 处理对应序号 1，B 处理对应序号 2……以此类推；再分配给每个受试对象一个两位随机数，并将随机数从小到大编序号 R，R 为对应试验单位所接受处理的序号，其所在的处理组便随机确定。

【例 4-11】 对比 A、B、C 3 种方法治疗某病的疗效，将入院时间相邻的 3 位受试者作为一个区组，共 10 个区组试分配处理。先将每个区组的 3 例按时间顺序从 1~3 编号，序号 1、2、3 分别对应处理 A、B、C，再给每一个受试者取一个 2 位随机数，并按区组别排序，按序号分配处理。方案见表 4-10。

表 4-10 10 个区组随机排列方案

区组 (入院时间)	随 机 数			随机数序号			随机排列结果		
	患 者 1	患 者 2	患 者 3	患 者 1	患 者 2	患 者 3	患 者 1	患 者 2	患 者 3
1	69	01	75	2	1	3	B	A	C
2	18	65	81	1	2	3	A	B	C
3	85	22	50	3	1	2	C	A	B
4	88	50	76	3	1	2	C	A	B
5	05	87	24	1	3	2	A	C	B
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10	90	64	16	3	2	1	C	B	A

如果采用 SAS 程序进行区组随机化，程序如下：

```
proc plan seed = 20121008;
output out = a;
factors block = 4 patients = 3;
run;
data b;
set a; no = _n_;
if patients = 1 then treat = 'A';
if patients = 2 then treat = 'B';
if patients = 3 then treat = 'C';
run;
proc print data = b;
var no block treat;
run;
```

程序说明：第一步，调用 PLAN 过程产生四个区组(每个区组有三个受试对象)的随机排列数，并把结果输出至数据集 a 中，选项 seed = ，规定产生随机数的种子。第二步，对数据集 a 进行操作，把患者按随机排列数进行分组，随机数等于 1 的进入 A 组，等于 2 的进入 B 组，等于 3 的进入 C 组，并将结果分别保存至数据集 b 中。最后通过 PRINT 过程将随机分组的结果展现出来。

SAS 运行结果如下：

Obs	no	block	treat
1	1	2	C
2	2	2	B
3	3	2	A
4	4	4	A
5	5	4	B
6	6	4	C
7	7	1	A
8	8	1	C
9	9	1	B
10	10	3	B
11	11	3	C
12	12	3	A

(3) 分层随机分组方法。

若样本含量大，可采用分段随机分组。其基本方法是将受试对象根据重要临床特征或危险因素(年龄、病情、性别等)分为多个阶段，每个阶段只对 m 个受试单位随机分组。 m 必须是处理数的整数倍，最好是 5 倍以上，然后利用完全随机分组的方法，对每段中的 m 个受试单位按其随机数字的大小由小到大排序，并按序号 R 分配处理。

【例 4-12】 试将 200 名肥胖患者随机等分为两组，比较两种减肥药的效果。

先将体重相近的 10 个受试对象分为 1 段，即 $S_{P(X>t)} = P(X > t) \times \sqrt{\sum \frac{1 - P(X > t)}{n - d}}$ ，规定每层随机排列序号 R 为 1~5 者为 A 处理组， R 为 6~10 者为 B 处理组，则需分 20 层完成，第 1、2 层的分组结果见表 4-11 及表 4-12。

表 4-11 分层随机分组举例(第 1 阶段)

受试者序号	第 1 阶段分组结果									
	1	2	3	4	5	6	7	8	9	10
随机数	421	333	459	418	384	400	213	391	754	318
<i>R</i>	8	3	9	7	4	6	1	5	10	2
处理	<i>B</i>	<i>A</i>	<i>B</i>	<i>B</i>	<i>A</i>	<i>B</i>	<i>A</i>	<i>A</i>	<i>B</i>	<i>A</i>

表 4-12 分层随机分组举例(第 2 阶段)

受试者序号	第 2 阶段分组结果									
	1	2	3	4	5	6	7	8	9	10
随机数	391	910	351	024	425	293	361	660	081	911
<i>R</i>	6	9	4	1	7	3	5	8	2	10
处理	<i>B</i>	<i>B</i>	<i>A</i>	<i>A</i>	<i>B</i>	<i>A</i>	<i>A</i>	<i>B</i>	<i>A</i>	<i>B</i>

如果采用 SAS 程序进行分层随机化，程序如下：

```
proc plan seed = 20121009;
output out = a;
factors period = 10 patients
= 20;
run;
data b;
set a; no = _n_;
if patients <= 10 then treat = 'A';
if 10 < patients <= 20 then treat = '
B';
run;
proc print data = b;
var no period treat;
run;
```

程序说明：第一步，调用 PLAN 过程产生 10 个分段(每个分段有 20 个受试对象)的随机排列数，并把结果输出至数据集 a 中，选项 seed = ，规定产生随机数的种子。第二步，对数据集 a 进行操作，把患者按随机排列数进行分组，随机数小于等于 10 的进入 A 组，大于 10 小于等于 20 的进入 B 组，并将结果分别保存至数据集 b 中。最后通过 PRINT 过程将随机分组的结果展现出来。

SAS 运行结果如下：

Obs	no	period	treat
1	1	6	B
2	2	6	A
3	3	6	B
.....			
198	198	8	A
199	199	8	A
200	200	8	B

(4)落实随机原则要贯穿于临床试验整个过程，除上面受试对象的分组随机化外，样本的抽样和受试对象的观察顺序都要遵循随机原则。

①病历的随机化抽取。

临床研究中有时还需从多年累积的某种患者病历中随机抽取若干份进行分析。可先把全部病历统一编号并按照事先确定的抽样比例划分为若干段，进行系统抽样。如计划按 1:20 比例抽样，则 1~20 号病历为第 1 段，21~40 号病历为第 2 段，依此类推，然后产生一系列随机数依次写在各段号下面，并用段内例数 20 去除，根据余数确定段内被抽中的病例数，见表 4-13。

表 4-13 病例的随机化抽取(按 1:20 比例)

病例分段号	1	2	3	4	5	6	7	8
段内病例编号	1~20	21~40	41~60	61~80	81~100	101~120	121~140	141~160
随机数字*	58	71	96	30	24	18	46	23
用 20 除所得余数	18	11	16	10	4	18	6	3
抽中的病例编号	18	31	56	70	84	118	126	143

注：* 代表来自随机数字表的第 10 行数据。

如果采用 SAS 程序按比例随机化抽样，程序如下：

```
data a ;
do i = 1 to 160 ;
output ;
end ;
run ;
proc surveyselect data=a method=sys samprate=0.05 seed=20141013 out=b;
run;
proc print data=b;
run;
```

程序说明：第一步，建立数据集 a，变量 i 表示 160 名学生的编号。第二步，对数据集 a 进行操作，使用的 sys 即系统抽样(systemic random sampling)，samprate 即抽样比例，也可采用 sampsize=8，可获得相同的结果，seed 规定产生随机数的种子。并将结果分别保存至数据集 b 中。最后通过 PRINT 过程将随机分组的结果展现出来。

SAS 运行结果如下：

Obs	i
1	11
2	31
3	51
4	71
5	91
6	111
7	131
8	151

②试验观测顺序的随机化。

当需要进行试验观测的受试对象人数较多时，随着观测时间的延长，所用观测仪器的工作状态、试剂的浓度、试验观测人员的心理和生理状态都可能产生一些变化，从而可能会给观测结果带入某些偏性。试验观测顺序的随机化的目的就是为了解决这一问题，通过随机安排试验顺序，保证对比各组中的每一受试对象都有同等机会处于任何序位上接受试验观测，从而可以在一定程度上抵消或控制时间顺序可能带来的偏性。具体方法是，先产生一系列随机数依次写在各受试者编号

下面(遇有数值相同的随机数舍去),然后按照随机数字的大小顺序(由小到大或由大到小)安排试验观测顺序。

3. 盲法

在科研设计中,研究对象和研究者的主观因素往往会影响到研究信息的真实性,产生信息偏倚。这种偏倚可产生于从设计到结果分析的任一环节。如在临床试验设计中,患者的心理因素和医务工作者的主观判断都可能干扰试验结果,产生偏倚。采用盲法设计可避免这种偏倚。所谓盲法设计(blinding design)是指通过临床试验设计使参与研究的受试者和(或)研究者、其他有关工作人员不知道受试对象接受的是试验措施还是对照措施,从而避免参与人员的心理效应所带来的主观偏性。如研究者对试验措施或阳性结果的偏爱,受试者对新药的过高期望或疑虑心理都会对观测结果产生影响;尤其是一些以受试对象的主观感觉或研究观测人员的主观判断为确定临床效应依据的临床试验,盲法设计尤为重要。根据盲法设置程度不同,一般可以分为非盲、单盲、双盲和三盲试验。

(1) 非盲法(open-label)。

非盲试验即不设盲的试验,研究者和受试者都知道具体治疗方案,如手术组与非手术组的比较。在非盲试验中,由于研究者或受试者对试验的信赖,或受试者对研究者的信任,在填写记录时某些受主观因素影响较大的指标就可能出现先入为主的观念。例如,当研究者知道受试者所接受的是试验药物时,可能对受试者的治疗情况倍加关心,如增加检查的频度,甚至护理人员也会格外关心该受试者,他们的这种行为很可能会影响受试者的态度,从而不知不觉影响观察指标的真实性。而当受试者知道自己所用的是安慰剂或对照药后,也会产生心理影响,妨碍或干扰与研究者在临床研究上的配合,造成偏倚。因此,即使在开放试验中,研究者或参与试验效应评价的人员最好不是同一个人。如果使参与评价的人员在评判过程中始终处于盲态,就能将变异程度控制到最低限度。非盲试验适用于有客观观察指标的试验,例如以客观的疾病或健康指标为评价效果的观察或改变生活习惯的干预措施等。另外,有些试验无法采用盲法观察,如评价某疾病手术、饮食和其他公共卫生措施的效果。

(2) 单盲法(single-blind)。

研究中只对研究对象设盲。如在新药试验中,患者不知道自己的分组情况和具体的用药情况,而参加试验的医护人员知道。该设计方法简单,易操作,可以消除受试者的心理因素影响,同时研究者可以较好地观察研究对象接受不同处理措施后的效果,及时处理治疗中可能发生的问题,保障研究对象的安全。缺点是在获得和分析各种研究资料如结局的判断时,可能受到研究者的主观因素影响而产生偏倚。例如,因为了解了研究对象的分组情况,医务人员在判断疗效时对治疗组和对照组病人采用不同的标准;或因为对照组没有得到治疗而有意无意地给对照组补偿性治疗等。实施单盲时,对照组需使用安慰剂。当安慰剂不利于患者病情时,可使用标准药物。

(3) 双盲法(double-blind)。

研究设计者安排和控制整个试验,研究对象和给予干预或结局评估的研究人员均不了解试验分组情况。双盲观察可以避免研究者和研究对象的主观因素影响带来的偏倚,提高研究的真实性。如在临床试验中患者和执行医疗措施的医务人员均不知道患者接受何种治疗,可使医生在检查、询问患者时一视同仁,不带主观偏见。该法的缺点是操作难度大,出现意外时很难及时进行处理,因此不适用于治疗过程中疗效变化大的试验和危重病人的治疗。

在药物临床试验中,为了落实“双盲”一定要使试验药与对照药的外形、颜色、气味、溶解度、味觉和外包装都高度一致。双盲设计要求有一整套完善的代号,如全部受试者、相关记录、使用的药物或安慰剂以及化验单都要使用代码。保密是双盲设计的关键,试验结束前,盲底泄露或退出试

验的研究对象超过 20%，则双盲试验失败。另外，应预先制定一些观察指标，以明确停药或更换药物的指征。在盲法实施过程中要注意避免出现违背医德的现象。当研究对象出现严重的副反应、治疗无效或病情加重时，应中止盲法，给予相应的处理。

(4) 三盲法 (triple-blind)。

研究对象、给予干预或结局评估的研究人员以及负责资料收集和分析的人员均不了解试验分组情况。该方法可避免资料分析人员带来的偏倚，设计更为合理，但执行难度大，常因医德、沾染、补偿和失密等问题而难以坚持。当受试对象的病情突然发生变化时，难以及时掌握治疗情况，造成处理延误。尽管三盲试验是减少偏倚最有效的方法，但由于其局限性使得实际工作中的应用并不普遍，最常应用的是“双盲”随机对照试验。

总的来说，盲法设计的关键是一定要做到“真盲”，而且，这种“真盲”要从始至终地保持在整个临床试验的全过程，直到在统计分析结束后才能揭盲，在此之前，由于任何原因导致的盲底泄露都称为破盲，一旦全部或大部分盲底泄露则意味着双盲试验的失败，需要重新安排另一次新的临床试验。

盲法观察适用于疗效评价依赖于患者或医生的药物试验，可根据研究情况具体选择方法。例如，主要根据患者的主诉来决定药效的试验可用单盲法；主要由医生主观判断决定药效的试验须用双盲法。当然，并不是所有的临床研究都需要或能够采用盲法设计。例如比较手术疗法与放射疗法治疗乳腺癌的效果，就不必、也无法采用盲法。在观察性研究中，盲法观察的原则要求我们在收集资料(疾病的结局或暴露史)时，对每个研究对象(病例或对照，暴露组或非暴露组)采用同样的方法收集资料、给予同等程度的重视；或者在资料收集时，调查员不知道所调查对象的分组情况，以客观地收集资料。

盲法设计较为复杂，执行起来也比较困难，应用时必须考虑其可行性并且要符合科学伦理道德原则，有一套保障受试者安全的有效措施，一旦发现紧急情况能及时对有关病例揭盲并采取相应对策。对病情复杂或危重病例一般不宜采用盲法设计。

4. 重复原则

重复 (replication) 是指在相同的试验条件下，进行多次研究或多次观察，主要包括三个方面的含义。

(1) 整个试验的重复。一个试验具有可重复性才具有可靠性、科学性。

(2) 同一试验单位的重复观察。这是为了保证观察结果的精度。例如，一般对血压测量 3 次，以 3 次的平均值为最终测量结果。

(3) 多个试验单位的重复观察。由于变异的存在、试验误差的产生，使同一试验条件下的试验单位具有不同的观察结果，为了避免把个别现象误认为普遍情况，必须在相同试验条件下，通过一定数量试验单位的重复观察，才能得出可靠的结论，掌握观察结果变异的规律，这里的“一定数量”就是指样本含量。一般来说，样本含量大，抽样误差就小，但样本含量过大时，将加大试验规模，延长试验时间，浪费人力物力，而且，样本含量过大还会增加控制试验观测条件的难度，有可能引入非随机误差，给观测结果带来偏性。因此，在试验设计中落实重复原则的一个重要问题就是如何科学、合理地确定样本含量。

4.2.3 科研设计中样本含量的确定

1. 影响样本含量的关键点

由于在各对比组例数相等时进行统计推断的效能最高，因此多数情况下都是按各组样本含量相等来估计。但在个别情况下，也可能要求各组样本含量按一定比例来估计。

在估计样本含量之前,首先要对以下几个与样本含量估计有关的统计学参数加以确定或做出估计。

(1) 规定有专业意义的差值 δ , 即所比较的两总体参数值相差多大以上才有专业意义。 δ 是根据试验目的人为规定的, 但必须有一定专业依据。 δ 值越小, 所需样本含量也越大。

(2) I 类错误(“弃真”的错误)的概率 α , 即当对比的双方总体参数值没有差到 δ 。但根据抽样观测结果错误地得出二者有差别的推断结论的可能性, α 确定得越小, 所需样本含量越大。在确定 α 时还要注意明确是单侧检验的 α , 还是双侧检验的 α 。在同样大小的 α 条件下; 双侧检验比单侧检验需要更大的样本含量。

(3) 检验效能, 用 $1 - \beta$ 表示。 β 为允许犯 II 类错误(“取伪”的错误)的概率。检验效能就是推断结论不犯 II 类错误的概率 $1 - \beta$, 称为把握度, 即当对比双方总体参数值间差值确实达到 δ 以上时, 根据抽样观测结果在规定的 α 水准上能正确地做出有差别的推断结论的可能性。在科研设计中, 常把 $1 - \beta$ 定为 0.90 或 0.80。一般来说, $1 - \beta$ 不宜低于 0.75, 否则可能出现非真实的阴性推断结论。

(4) 总体标准差 σ 或总体率 π 的估计值。它们分别反映计量数据和计数数据的变异程度。一般是根据前人经验或文献报道做出估计。如果没有前人经验或文献报道作为依据, 可通过预试验取得样本的标准差 s 或样本率 P 分别作为 σ 和 π 的估计值。 σ 的估计值越大, π 的估计值越接近 0.5, 所需样本含量越大。

(5) 资料的性质。一般来说, 在同等条件下, 数值变量所需样本含量小于分类变量。

(6) 设计的类型。由于在同样的要求和条件下, 完全随机设计(成组设计)所需样本含量最大, 故一般都要按完全随机设计做出样本含量的估计。

(7) 比较类型。在成组设计时, 还可根据假设检验中具体采用什么样的比较类型, 分为差异性检验、非劣效性检验、等效性检验和优效性检验, 对应的样本含量估计公式有所不同。

2. 常用样本含量估算方法

(1) 经验法: 难治的疾病、疗效有差别时(如癌症)需 5 ~ 10 例; 急重病死亡率高的(如休克、心衰、肾衰、呼吸衰竭)需 30 ~ 50 例; 一般病和慢性病需 100 ~ 300 例等。

(2) 算法: 根据影响样本含量的各因素的要求, 以及差别假设检验的类型和资料的类型, 具体的计算公式如下。

① 样本均数与总体均数比较(或配对比较)。

用式(4-1)计算样本含量 n 。

$$n = \left[\frac{z_{1-\alpha} + z_{1-\beta}}{\delta} \right]^2 \quad (4-1)$$

【例 4-13】研究新药提升白细胞的疗效, 由预试验得出用药前后白细胞差值的标准差为 1.5 千个/ mm^3 , 当白细胞平均上升 1 千个/ mm^3 时认为临床有效, 临床试验需要多少人?

【解】本例 $\delta = 1$, $\sigma = 1.5$, 规定单侧 $\alpha = 0.05$, $\beta = 0.10$, $Z_{1-0.05} = 1.645$, $Z_{1-0.10} = 1.282$, 代入式(4-1),

$$n_1 = \left[\frac{(1.645 + 1.282) \times 1.5}{1} \right]^2 = 19.3, \text{取 } 20。$$

故要满足上述条件, 需要样本量为 20 人。

配对设计时, 用式(4-1)计算得的 n 值为对子数。公式中的 χ^2 配对数据差值的标准差。

本例 SAS 程序如下:

```
proc power;
onesamplemeans
```

```
mean =1
stddev =1.5
ntotal =.
sides =1
alpha =0.05
power =0.9;
run;
```

程序说明：在 SAS 中的 POWER 过程中选择语句 ONE SAMPLE MEANS,进行单样本均数与总体均数的比较；后面的各语句用于计算样本量时需要的参数赋值，包括 MEAN 样本均数；STDDEV 样本标准差，NTOTAL 总例数，如需要计算样本量则其赋值为缺失，以“.”表示，SIDES 定义为单侧检验（赋值为 1），还是双侧检验（赋值为 2）；ALPHA 检验水准；POWER 检验效能。

SAS 运行结果如下：

Computed N	Total
Actual	N
Power	Total
0.904	21

②两样本均数比较。

n_1 、 n_2 为两样本含量，一般要求相等，用式(4-2)计算：

$$n_1 = n_2 = 2 \left[\frac{z_{1-\alpha} + z_{1-\beta}}{\delta} \right]^2 \tag{4-2}$$

【例 4-14】 某医师研究一种降低高血脂患者胆固醇药物的临床疗效，以安慰剂作为对照。事前规定试验组与对照组相比，平均多降低 0.5mmol/L 以上，才有推广应用价值。而且由有关文献中查到高血脂患者胆固醇值的标准差为 0.8mmol/L，若要求犯 I 类错误的的概率不超过 5%，犯 II 类错误的概率不超过 10%，且要两组例数相等则每组各需观察多少例？

【解】 本例 $\delta = 0.5\text{mmol/L}$ ， $\sigma = 0.8\text{mmol/L}$ ，规定单侧 $\alpha = 0.05$ ， $\beta = 0.10$ ， $Z_{1-0.05} = 1.645$ ， $Z_{1-0.10} = 1.282$ ，代公式(4-2)，

$$n_1 = n_2 = 2 \times \left[\frac{(1.645 + 1.282) \times 0.8}{0.5} \right]^2 = 43.86, \text{取 } 44。$$

故要达到上述要求，两组至少各需观察 44 例。

本例 SAS 程序如下：

```
proc power;
twosamplemeans
test =diff
meandiff =0.5
stddev =0.8
npergroup =.
sides =1
alpha =0.05
power =0.9;
run;
```

程序说明：在 SAS 中的 POWER 过程中选择语句 TWO SAMPLE MEANS,进行两样本均数的比较；后面的各语句用于计算样本量时需要的参数赋值，包括 MEANDIFF 容许误差；STDDEV 两样本合并标准差，NPERGROUP 为每个样本例数，如需要计算样本量则其赋值为缺失，以“.”表示，

SIDES 定义为单侧检验(赋值为 1), 还是双侧检验(赋值为 2); ALPHA 检验水准; POWER 检验效能。

SAS 运行结果如下:

Computed	N Per Group
Actual	N Per
Power	Group
0.903	45

③多个样本均数比较。

每组样本含量为 n , 其估计值用式(4-3)计算。

$$n = \psi^2 (\sum S_i^2/k) / \lfloor \sum (\bar{X}_i - \bar{X})^2 / (k - 1) \rfloor \tag{4-3}$$

式中, k 为组数, \bar{X}_i 、 S_i 分别为各组的均数与标准差的估计值, $\bar{X} = \sum \bar{X}_i/k$, ψ 为界值, 由专门的 ψ 界值表(因篇幅所限, 本书中未将此表附在书后, 现在几乎都是通过统计软件来实现计算, 不需要用户去查表了)查得。计算时先用自由度 $v_1 = k - 1$, $v_2 = \infty$ 时的 ψ 代入式中求 n_1 , 再用 $v_1 = k - 1$, $v_2 = k(n - 1)$ 的 ψ 代入式中求 n_2 , 第三次用 $v_1 = k - 1$, $v_2 = k(n_2 - 1)$ 的 ψ 代入式中求 n_3 , 依次类推, 直至前后两次求得的结果趋于稳定为止。

【例 4-15】 研究三种药物的退热效果, 经预试验退烧的平均天数分别是 2.67、11.67、6.83, 标准差分别是 3.67、10.63、6.27。设 $\alpha = 0.05$, $\beta = 0.10$ 。问: 每组需观察多少病例?

【解】 本例 $\bar{X} = (2.67 + 11.67 + 6.83)/3 = 7.06$

$$\sum S_i^2 = 3.67^2 + 10.63^2 + 6.27^2 = 165.78$$

$$\sum (\bar{X}_i - \bar{X})^2 = (2.67 - 7.06)^2 + (11.67 - 7.06)^2 + (6.83 - 7.06)^2 = 40.58$$

$k = 3$, 以 $v_1 = 3 - 2 = 2$, $v_2 = \infty$, 查 ψ 界值表, 得 $\psi = 2.52$, 代入公式,

$$n_1 = 2.52^2 \times (165.78/3) / [40.58/(3 - 1)] = 17.3, \text{取 } 18。$$

以 $v_1 = 3 - 1 = 2$, $v_2 = 3 \times (18 - 1) = 51$, 查 ψ 界值表, 得 $\psi_{2,51} \approx 2.59$, 再代入公式计算,

$$n_1 = 2.59^2 \times (165.78/3) / [40.58/(3 - 1)] = 18.3, \text{取 } 19。$$

两次计算结果接近, 故每组可观察 19 例, 三组共需 57 例。

本例 SAS 程序如下:

```
proc power;
onewayanova
groupmeans = 2.67 | 11.67 | 6.83
stddev = 7.43
npergroup = .
alpha = 0.05
power = 0.9;
run;
```

程序说明: 在 SAS 中的 POWER 过程中选择语句 ONE WAY ANOVA, 进行多组样本均数的比较; 后面的各语句用于计算样本量时需要的参数赋值, 包 GROUP MEANS 各组样本均数, 以“|”为分隔; STDDEV 样本标准差, 可用三组合并标准差, 保守做法是采用三组中最大标准差, 本例采用合并标准差 $\sqrt{\frac{2.67^2 + 11.67^2 + 6.83^2}{3}}$, NTOTAL 总例数, 如需要计算样本量则其赋值为缺失, 以“.”表示, SIDES 定义为单侧检验(赋值为 1), 还是双侧检验(赋值为 2); ALPHA 检验水准; POWER 检验效能。

SAS 运行结果如下：

Computed	N Per Group
Actual	N Per
Power	Group
0.912	19

④样本率与总体率比较(大样本)。

样本含量 n 用式(4-4)计算。

$$n = \pi_0 \times (1 - \pi_0) \times \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\delta} \right)^2 \tag{4-4}$$

式中, π_0 为已知总体率, $\delta = \pi_1 - \pi_0$, π_1 为预期试验结果的总体率。

【例 4-16】 研究新药治疗某病的疗效, 预计有效率是 90%, 已知老药的有效率是 80%, 给定双侧 $\alpha = 0.05$, $\beta = 0.10$ 。问: 至少要观察多少病例?

【解】 本例 $\pi_0 = 0.80$, $\pi_1 = 0.90$, $\delta = 0.90 - 0.80 = 0.10$, 双侧 $\alpha = 0.05$, 则 $Z_{1-0.05} = 1.645$, $Z_{1-0.10} = 1.282$, 代入式(4-4)。

$$n = 0.80 \times (1 - 0.80) \times \left(\frac{1.96 + 1.282}{0.10} \right)^2 \approx 169$$

本试验需观察 169 个病例。

本例 SAS 程序如下：

```
proc power;
onesamplefreq test=z
nullproportion=0.8
proportion=0.9
ntotal=.
method=normal
sides=2
alpha=0.05
power=0.9;
run;
```

程序说明: 在 SAS 中的 POWER 过程中选择语句 ONE SAMPLE FREQ TEST = z, 进行单样本率数与总体率的比较。需注意, 在样本率与总体率比较的样本含量估计中, 必须通过选项 TEST 定义检验方法。本例采用正态近似法, 还可以根据实际情况选用确切概率法, TEST 的赋值为“exact”; 后面的各语句用于计算样本量时需要的参数赋值, 包括 NULL PROPOTION 总体率; PROPOTION 样本率, NTOTAL 总例数, 如需要计算样本量则其赋值为缺失, 以“.”表示, SIDES 定义为单侧检验(赋值为 1), 还是双侧检验(赋值为 2); ALPHA 检验水准; POWER 检验效能。

SAS 运行结果如下：

Computed	N Total
Actual	N
Power	Total
0.901	137

⑤两个样本率比较。

当样本率服从二项分布时, 若样本率介于 0.2 ~ 0.8 之间, 按正态和近似正态分布原理, 其样

本含量计算公式为：

$$n_1 = n_2 = 2 \frac{(z_{1-\alpha} + z_{1-\beta})^2 p(1-p)}{(p_1 - p_2)^2} \tag{4-5}$$

n_1 、 n_2 为两样本含量，一般取其相等，用式(4-5)计算，式中 p_1 、 p_2 分别为两总体率的估计值， $p = (p_1 + p_2)/2$ 。当率偏离上述范围时，对率进行平方根反正弦变换，以“弧度”为单位计算，见式(4-6)。

$$n_1 = n_2 = \frac{1}{2} \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\arcsin \sqrt{p_1} - \arcsin \sqrt{p_2}} \right)^2 \tag{4-6}$$

【例 4-17】 比较甲、乙两种吸入剂对哮喘的缓解情况，由以往资料知甲、乙两种吸入剂对哮喘的缓解率为 45% 和 25%。问：每组最少需观察多少病例？

【解】 本例 $p_1 = 0.45$ ， $p_2 = 0.25$ ，取双侧 $\alpha = 0.05$ ，则 $z_{1-0.05} = 1.96$ ， $\beta = 0.10$ ， $z_{1-0.10} = 1.282$ ，代入式(4-5)， $p = (0.45 + 0.25)/2 = 0.35$ 。

$$n_1 = n_2 = 2 \frac{(1.96 + 1.282)^2 0.35(1 - 0.35)}{(0.45 - 0.25)^2} = 59.78 \times 2 \approx 120$$

每组需观察 120 例，两组共需 240 例。本例 SAS 程序如下：

```
proc power;
  twosamplefreq
  groupproptions=0.45|0.25
  nullproportiondiff=0
  npergroup=.
  sides=2
  alpha=0.05
  power=0.9;
run;
```

程序说明：在 SAS 中的 POWER 过程中选择语句 TWO SAMPLE FREQ TEST，进行两样本率的比较，后面的各语句用于计算样本量时需要的参数赋值，包括 GROU PPROPOTIONS 两组样本率，以“|”隔开；NULLPROPOTIONDIFF 检验假设中两总体率的差值，NPERGROUP 每个组的样本量，如需要计算样本量则其赋值为缺失，以“.”表示，SIDES 定义为单侧检验（赋值为 1），还是双侧检验（赋值为 2）；ALPHA 检验水准；POWER 检验效能。

SAS 运行结果如下：

Computed	N Per Group
Actual	N Per
Power	Group
0.900	120

⑥多个样本率比较。

每组样本含量为 n ，其值用式(4-7)计算。

$$n = \frac{2\lambda}{(2\arcsin \sqrt{p_{\max}} - 2\arcsin \sqrt{p_{\min}})^2} \tag{4-7}$$

式中， p_{\max} 、 p_{\min} 分别为最大率与最小率； λ 是界值，当给定 α 、 β 和自由度 $v = k - 1$ 时，由附表 3 查得， k 为组数。

【例 4-18】 研究三种方法治疗慢性气管炎的近控率，初步试验得甲法近控率为 40%，乙法为 55%，丙法为 65%。设 $\alpha = 0.05$ ， $\beta = 0.10$ 。问：共需要观察多少病例？

【解】 本例 $p_{\max} = 0.65$, $p_{\min} = 0.40$, $v = k - 1 = 2$, $\alpha = 0.05$, $\beta = 0.10$, 由附表 3 查得 $\lambda = 12.65$, 代入式(3-6) ,

$$n = \frac{2 \times 12.65}{(2\arcsin \sqrt{0.65} - 2\arcsin \sqrt{0.40})^2} \approx 99$$

每组需观察 99 例 , 三组共需 297 例。

本例 SAS 程序如下 :

```
data nkkysj4_2_18;
alpha=0.05;
beta=0.10;
df=2;
pmax=0.65;
pmin=0.40;
ars1=arsin(sqrt(pmax));
ars2=arsin(sqrt(pmin));
chi=cinv(1-alpha,df);
lam=cnonct(chi,df,beta);
n_per_group=lam/(2*(ars1-ars2)
** 2);
proc print;
var n_per_group;
run;
```

程序说明 : SAS 未提供专门的过程直接计算多组样本率比较问题的样本量。根据其计算公式 , 上述程序中各语句包括 : ALPHA 检验水准 , BETA 是第 II 类错误的概率 , 即 $1 - \text{POWER}$; DF 自由度 , $DF = g - 1$, g 为组别数 ; PMAX 各组样本率的最大值 ; PMIN 各组样本率的最小值 ; ARS1、ARS2 分别计算 PMAX、PMIN 平方根的反正弦值 ; CHI 为检验水准 ALPHA 下 , 自由度为 df 时 , 卡方分布的界值 ; LAM 为中心卡方分布的非中心参数 , 通过函数 CNONCT 计算 ; N_PER_GROUP 为每个组的样本量。采用 PRINT 过程输出结果到 OUTPUT 窗口 , 用 VAR 语句定义仅输出各组样本量的估计值 N_PER_GROUP。

SAS 运行结果如下 :

	n_per_
Obs	group
1	98.8252

差别检验时采用上面提及的公式进行样本含量的计算。除差别检验外 , 临床试验中还有优效性、等效性和非劣效性检验。当研究者认为试验药比对照药优于某一具有临床意义的数值 Δ 时才为优效 , 此时采用临床优效性检验 (Δ 为优效性界值)。检验一种药物是否与另一种药物的疗效“相等” (实际为相差不超过一个指定的数值 Δ , 即等效性界值) , 称为等效性试验。非劣效性试验为检验一种药物是否不劣于另一种药物的试验 , 其设计要求阳性对照药物应具有稳定的有效性 , 一般用于有客观疗效指标的临床研究中 (如抗菌药物的临床终点、肿瘤治疗中死亡或进展事件等)。试验药总体疗效比对照药的总体疗效要差 , 且两药总体疗效之差大于或等于 Δ (Δ 为非劣效性界值)。优效性、等效性和非劣效性检验根据资料类型不同 , 样本含量的计算公式不同。两样本率比较时 , 优效性、等效性和非劣效性检验样本含量计算公式如下。

①优效性检验样本含量 :

$$n_c = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{(\pi_T - \pi_c - \Delta)^2} \left[\frac{\pi_T(1 - \pi_T)}{K} + \pi_c(1 - \pi_c) \right] \tag{4-8}$$

$$n_T = Kn_c \quad (4-9)$$

n_c 为对照组样本含量。 n_T 为试验组样本含量。公式中 π_T 为试验组率, π_c 为对照组率, Δ 为优效界值, α 为第 I 类错误 (常取单侧 0.025), β 为第 II 类错误 (常取单侧 0.20), K 为试验组与对照组例数分配比。

②等效性试验样本含量:

$$n_c = \frac{(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{(\Delta - |\pi_T - \pi_c|)^2} \left[\frac{\pi_T(1 - \pi_T)}{K} + \pi_c(1 - \pi_c) \right] \quad (4-10)$$

n_c 为对照组样本含量, 公式中 π_T 为试验组率, π_c 为对照组率, Δ 为等效性界值, α 为第 I 类错误 (常取双侧 0.10), β 为第 II 类错误 (常取双侧 0.20), K 为试验组与对照组例数分配比; n_T 采用式 (4-9) 计算。

③非劣效性检验样本含量:

$$n_c = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{(\pi_T - \pi_c + \Delta)^2} \left[\frac{\pi_T(1 - \pi_T)}{K} + \pi_c(1 - \pi_c) \right] \quad (4-11)$$

n_c 为对照组样本含量。公式中 π_T 为试验组率, π_c 为对照组率, Δ 为非劣效性界值, α 为第 I 类错误 (常取单侧 0.025), β 为第 II 类错误 (常取单侧 0.20), K 为试验组与对照组例数分配比; n_T 采用式 (4-9) 计算。

两样本均数比较的优效性、等效性和非劣效性检验样本含量计算公式如下。

①优效性检验样本含量:

$$n_c = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2 \sigma^2 \left(1 + \frac{1}{K}\right)}{(\mu_T - \mu_c - \Delta)^2} \quad (4-12)$$

n_c 为对照组样本含量。公式中 μ_T 为试验组均数, μ_c 为对照组均数, Δ 为优效性界值, σ 为标准差 (假设两组标准差相同), α 为第 I 类错误 (常取单侧 0.025), β 为第 II 类错误 (常取单侧 0.20), K 为试验组与对照组例数分配比; n_T 采用式 (4-9) 计算。

②等效性试验样本含量:

$$n_c = \frac{(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2 \sigma^2 \left(1 + \frac{1}{K}\right)}{(\Delta - |\mu_T - \mu_c|)^2} \quad (4-13)$$

n_c 为对照组样本含量。公式中 μ_T 为试验组均数, μ_c 为对照组均数, Δ 为等效性界值, σ 为标准差 (假设两组标准差相同), α 为第 I 类错误 (常取双侧 0.10), β 为第 II 类错误 (常取双侧 0.20), K 为试验组与对照组例数分配比; n_T 采用式 (4-9) 计算。

③非劣效性检验样本含量:

$$n_c = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2 \sigma^2 \left(1 + \frac{1}{K}\right)}{(\mu_T - \mu_c + \Delta)^2} \quad (4-14)$$

n_c 为对照组样本含量。公式中 μ_T 为试验组均数, μ_c 为对照组均数, Δ 为非劣效性界值, σ 为标准差 (假设两组标准差相同), α 为第 I 类错误 (常取单侧 0.025), β 为第 II 类错误 (常取单侧 0.20), K 为试验组与对照组例数分配比; n_T 采用式 (4-9) 计算。

4.2.4 内科学试验设计中实例展示

1. 随机原则的实例展示

【例 4-19】 观察小剂量干扰素加三氮唑核苷治疗流行性乙型脑炎的疗效的研究。在一般治

疗的基础上加用小剂量干扰素及三氮唑核苷治疗流行性乙型脑炎 99 例,采用同期的、接受一般治疗的 73 例该病患者作为对照。治疗组:轻型 29 例,普通型 40 例,重型 22 例,极重型 8 例;对照组:73 例,轻型 18 例,普通型 32 例,重型 17 例,极重型 6 例。两组病人均采用传统降温、镇静、降颅内压、肾上腺皮质激素及抗生素预防感染等对症治疗。在此基础上,治疗组选择发病在 5d 以内的病人,加用干扰素和三氮唑核苷静滴,疗程为 5~7d。两组比较疗效,差别有统计学意义,结论是在一般治疗的基础上加用小剂量干扰素及三氮唑核苷治疗流行性乙型脑炎的疗效优于一般治疗的疗效。

【实例辨析】 本例中,研究者考虑到病情的严重程度可能是影响疗效的一个影响因素,因此对两组中病情的严重程度做了安排,经统计学检验,两组病情严重程度的构成上差别没有统计学意义,好像具有可比性。但研究者人为选择发病在 5d 以内的病人进入治疗组,而对照组没有此限制。正确的做法:对照组也应选择发病 5d 以内的病人,在试验分组时遵循随机的原则,以使病人有相同的机会进入治疗组和对照组。在样本含量不是很大的情况下,按照病情等重要非试验因素进行分层随机化,可能更为合适。

2. 对照原则的实例展示

【例 4-20】 观察某种药物对合并糖尿病的急性脑梗死患者治疗的临床效果。收集 2008 年 1 月至 2009 年 12 月的 63 例合并 II 型糖尿病的急性脑梗死患者,采用该药物治疗,与过去一年收治的 35 例合并 II 型糖尿病的急性脑梗死患者采用常规治疗的疗效进行比较。

【实例辨析】 本例中采用的是历史对照,与过去采用常规治疗的患者进行比较。历史对照的缺点:治疗组和对照组间除了研究药物外,有许多因素是不相似的,包括人口统计学特征、诊断标准、疾病的阶段及持续时间、伴随治疗以及观察的情况(如评价结果的方法、研究者的期望),很可能影响结果。因此正确的做法是,将同期合并糖尿病的急性脑梗死患者随机分为试验组和对照组,一组给予新药物,另一组给予常规治疗,再比较两种的疗效。

3. 盲法原则的实例展示

【例 4-21】 评价阿糖腺苷治疗单纯疱疹性脑炎的疗效,同时观察阿糖腺苷可能存在的药物不良反应。在美国某医疗中心住院的单纯疱疹性脑炎病人,通过简单随机分组的方法,分为 A 组和 B 组。对所有进入试验的病人进行脑穿术后进行治疗试验。脑穿疱疹病毒阳性者疗程为 10d;脑穿疱疹病毒阴性者为避免药物潜在的毒性作用,疗程为 5d。A 组用阿糖腺苷(A 制剂),B 组用阿糖腺苷(B 制剂)。两组阿糖腺苷剂量均为每日 15mg/kg 体重。阿糖腺苷(A 制剂)、阿糖腺苷(B 制剂)分别为安慰剂和阿糖腺苷,即用 A 制剂的病人使用的是安慰剂,他们是对照组;用 B 制剂的病人使用的是真正阿糖腺苷,他们是试验组。但并不将这一情况告知主治医生和病人。以保证双盲法的实施。主治医师负责填写阿糖腺苷治疗疱疹性脑炎疗效观察记录表。安慰剂和阿糖腺苷由同一厂家提供。

【实例辨析】 本例中,两组病人在年龄、性别、脑穿时的麻醉时间长短和脑穿的大小等方面都均衡可比。双盲的实施避免病人和主治医师的心理因素对试验效应可能造成的影响。以病死率为试验效应指标,阿糖腺苷可使脑穿证实的疱疹病毒阳性病人的病死率明显降低,而脑穿阴性的病采用人 A 和 B 制剂后的病死率没有统计学差别。头痛,瘙痒等不良反应在两组之间没有差别。

4. 重复原则的实例展示

【例 4-22】 为探讨脑血栓形成与缺铁性贫血之间的关系、临床表现及治疗。某研究回顾性分析 4 例缺铁性贫血并发脑梗死患者的临床资料。4 例患者均为中重度贫血,血小板增多,血压偏低,2 例存在慢性感染病灶。所有患者均有颈内动脉系统脑梗死,其中 2 例为分水岭梗死,2 例为

基底节梗死。吸氧、纠正贫血、调整血压在适当水平以及其他常规治疗有效。该研究者认为对于中重度缺铁性贫血患者，如伴有血小板增多和血压偏低等情况，易导致脑血栓形成。在常规治疗脑血栓形成的同时应注意寻找贫血原因并且纠正贫血。

【实例辨析】 对错的辨析与释疑：生物医学研究的现象常见变异性，只有在相同试验条件下进行多次独立重复试验，随机现象的变化规律性才能正确地显露出来。在该研究中，研究者只选取了4例患者，观察到的结果并不能体现共性，其相关结论不能随意外推。

4.3 临床试验的质量控制

在试验研究中，由于受试对象的情况复杂以及各种主、客观因素的干扰，使试验结果与其真值之间不完全一致，这种差别通常称为试验误差。试验误差的存在使研究者对试验结果的评价具有某种不确定性，因此，搞清楚误差的来源、控制试验误差是提高试验有效率、保证试验结果正确性的手段之一。试验误差分为三类：随机误差、系统误差和过失误差。

4.3.1 试验误差的种类

1. 随机误差

随机误差(random error)，又称偶尔误差，是由一系列偶然因素引起的不易控制的误差，其值具有统计规律性，如重复误差。在测量过程中，参与的变异成分越多，则测定值越分散，即随机误差越大。例如，同一名医生多次测量同一病人的血压值所产生的随机误差，要小于多名医生同时测量同一病人血压值所产生的随机误差。随机误差是不能避免的，但有其规律可循，大量观察表明，随机误差呈现以0为中心的正态分布。

随机抽样产生的抽样误差服从随机误差分布的规律。抽样误差越大，对试验结果进行统计推断的困难也就越大。试验设计的任务之一，就是要控制和减小随机误差，便于进行统计推断。

2. 系统误差

系统误差(systematic error)又称偏倚(bias)，是由于对受试对象、试验因素或条件等控制不严密而产生的一种误差，它使试验结果系统地偏离其值。根据来源不同，可将系统误差分为以下几种。

(1) 选择偏倚。

选择偏倚(selection bias)是由于对受试对象选择或分组不当所产生的偏倚。按照抽样理论，每个受试对象均应由随机抽样、随机分组得到，如果不严格遵守这一原则，则可能出现结果偏差。例如，在选择病例时未严格执行入选标准和剔除标准，或按照病人意愿进行分组等，都可能使试验结果产生偏倚。

(2) 信息偏倚。

信息偏倚(information bias)是由于测量仪器、操作方法、试验条件以及主观因素等方面的原因造成的偏倚。如某医生在观察某种手术的效果时，抱有非常乐观的态度，一旦出现负性结果，就倾向于找各种原因来解释。

(3) 混杂偏倚。

混杂偏倚(confounding bias)由影响试验结果的非试验因素在各对比组中不平衡所引起的偏倚。如疾病的转归除了药物的治疗作用外，还与病人的病情、体质等多方面的因素有关。若忽略了这些有关因素在各对比组的均衡性，就会发生混杂偏倚，导致错误的结论。

3. 过失误差

过失误差是由于在操作、登记、汇总、计算过程中过失造成的误差。这种误差所产生的影响往往比其他的误差大得多，必须严格防止，一旦发现，应立即纠正，无法纠正者则予以剔除。

4.3.2 误差的控制

试验设计、试验实施和结果分析的各个阶段都存在误差。因此，在多个阶段都要十分注意误差的控制，要针对误差产生的原因采取相应的控制措施，力求使研究得到接近真实的可靠结果。

1. 消除过失误差的方法

- (1) 必须有完整的原始记录。
- (2) 建立严格的核对检查制度。
- (3) 制定统一的操作规程。
- (4) 要有实事求是的科学作风。

2. 纠正系统误差的方法

- (1) 定期地或经常地校准仪器；试验因素标准化；试验过程的统一、规范、完整记录；参与研究的试验人员技术水平和医德水平的提高培训；盲法的应用等可以防止信息偏倚导致的系统误差。
- (2) 对研究总体的规定明确具体；正确拟定观察对象的纳入标准和排除标准；合适的样本含量，随机抽样、随机分组等防止选择性偏倚导致的系统误差。
- (3) 设计时明确试验因素和对试验效应有影响的非试验因素，以选用合适的试验方法，均衡非试验因素的影响；在试验阶段始终贯彻标准化的试验条件；在统计分析阶段采用合适的统计方法，以排除混杂偏倚导致的系统误差。

3. 减少随机误差方法

- (1) 要思想集中、工作细致，严格按照操作规程进行操作。
- (2) 必要时重复测定取均值。
- (3) 确定合适的样本含量。

4.4 随机双盲临床试验的实施过程

在药物临床试验中，多中心、随机、双盲临床试验是采用最多的一种。双盲临床试验需要主要研究者会同药品监督管理人员、统计学家，制定严格的操作规范，防止受试者分组编码不必要的扩散。在临床试验进行过程中，一旦泄盲，即受试者的分组情况被参与试验的相关人员知晓，试验将被视为无效，需要重新试验。

4.4.1 随机双盲临床试验实施过程

1. 随机化分组编码的产生

随机化分组编码，又称为盲底，是采用随机化方法写出的受试者分组（试验组或对照组）的随机安排。药物按盲底进行分配包装后，运往各试验中心。生成盲底和实施编盲的人员应独立于临床试验。盲底一般采用文件形式密封保存。不同的临床试验设计方案的随机化方法不同。随机数字的产生应该使用统计软件在电子计算机中产生，必须具有重现性。随机数字的产生时间距离药物编码分装的时间越近越好，一般为 2 天之内。

在临床试验之前,对受试对象进行编号,通常采用分层随机化方法来产生随机数字,基于随机数字将对应编号的受试对象分配进入相应的对比组中去。对于多中心临床试验,中心通常作为1个分层因素。当某些因素,如疾病的不同亚型等,会对治疗结果有影响时,也应将其作为分层因素。如疾病的不同亚型有3种,有3个中心参加了临床试验,则总共为9层。分层可以较好地保证层内的均衡性,但病例数过少时,不宜分层过多,否则分层后各个亚组的病例数将很少。在确定了样本大小、分层因素后,就可以在电子计算机上使用统计软件产生随机数,注意定义相应的种子数,以便于随机数的重现。分层随机化方法请参见第3章的相关内容。

2. 试验药物的准备

申办者需按试验研究设计方案准备每个受试者所使用的药物。假设试验药物为片剂(10mg/片),每天服用3次,每次2片,疗程为28天,方案中规定受试者每周随访一次,发放药物,将每周的用量放入一个瓶子中,共42片。实际操作中,需在每瓶中多放入8~9片的补充量,以备病人不慎将药片遗落毁损,或者不能满足下一次随访时之需,据此也可以估计病人的依从性;多余的药物需在下次随访时收回、清点,并在CRF(病例报告表)上做出记录。多余的药物需放回该病例的药盒中,直到全部试验结束,总结报告完成后,无异议时才可在监视下销毁。

如果每个病人需服药4周(28天),每周1瓶,则共需4瓶,如50片1瓶,则每个病人共需200片药。申办者需分别根据受试者的人数准备好试验组和对照组的所有用药。在分装药物时需有书面记录,例如共计多少药片,分装入多少瓶(每瓶多少片)后还剩余多少,必须清点无误。每瓶的装量需要复核,保证瓶中确为50片。装试验药与对照药的瓶子外形应相同。

药物装瓶后,需要在药瓶上贴上相应的标签,标签上主要标明“×××药物的临床研究”用药、内含多少片、用法、批号、生产厂家,还需设置编号栏,供编盲时填写或贴标签。如每个病人服药4瓶,则每4瓶应有一个大包装(药盒),该药盒需有标签,内容同上。对于有些临床试验,每次访视发药量有所不同时,还需在药瓶标签上标明第几次访视时发放。申办者应当准备足够的标签及药盒。

3. 应急信件的准备

为了保证受试者的安全,在双盲临床试验中,申办者需为每个受试者准备一个应急信件。通常使用纸质坚韧不透光的厚牛皮纸制成应急信封,信封上印有“×××药物的临床研究应急信件”、编号。编号必须由同一个人填写或为贴的标签,非必要时不得拆阅,试验结束时随CRF表一起收回、信封上还印有遇有下列情况由研究者决定是否拆阅:①严重不良事件;②病人需紧急抢救,如果拆阅,需注明拆阅者、拆阅日期、原因等。

信封内的信纸上印有×××药物临床研究、编号、受试者已被分入哪个处理组等信息,均由编盲者填写,信件密封后将随药物一同发往各个中心。

也可以将应急信件制作成更为方便的刮卡,卡片上打印上应急信件所要求的信息,将分组情况用特殊涂层材料遮盖住,不透光,保证在不刮开的情况下,无法看到分组情况。

应急信件(或应急卡)仅在病人发生严重不良事件、需紧急抢救,必须知道其分组情况时才能打开(或刮开)。

4. 药物分配编盲

申办者将药物、标签、药盒、应急信封及信纸准备就绪,统计学家在电子计算机上产生盲底后,即可进行药物分配编盲。药物分配编盲需在药品监督管理部门有关工作人员的监视下进行,其具体实施步骤如下所述。

(1) 试验药物抽样。

编盲前,先由药品检验所的工作人员在申办者所准备的药物中(试验组与对照组)进行随机抽

样，对试验药和对照药做出药检报告，内容包括外观、形状、大小、颜色等是否一样；鉴别两者的真伪；检查分解度、重量两者有无差异；测定是否含有有效成分等。

(2) 药物的包装。

统计学家根据已形成的盲底将药物分装入每个受试者所使用的药盒中，写上相应的编号，药盒内的小瓶上也写上该编号，当各次访视所发放药物不同时，还需填上第几次访视时发放。

(3) 应急信件。

统计学家根据盲底，填写应急信封及信纸，并密封。

5. 盲底的保存

盲底通常包括第一次揭盲与第二次揭盲两部分：第一次揭盲的盲底包含了所有病人的分组情况(如 A 组或 B 组、种子数、分层因素等信息)；第二次揭盲的盲底则标明了各组服用的是试验药还是对照药(如 A 组为试验药，B 组为安慰剂)。通常，先将第二次揭盲的盲底单独装入小信封后，再连同第一次揭盲的盲底装入大信封，进行保存。

盲底一式两份，密封后分别由申办者和参与研究的组长单位的临床药理基地保存。因为整个双盲临床试验的盲底全部在此信封内，一旦装在一个信封中的两份文件同时发生意外，就只能拆阅全部的应急信件才能获知，所以盲底分装成两份是为了防止意外。

6. 揭盲

双盲临床试验通常采用二次揭盲的方法，当所有受试者病例报告表中的数据全部录入计算机，并经过盲态审核后，数据将被锁定。此时进行第一次揭盲，将各病例号所隶属的组别(如 A 组与 B 组)的盲底，即分药盲底，告知生物统计学家，以便对数据进行统计分析。此时，仅仅知道各受试对象所处的不同分组，而不知道哪个为试验组，哪个为对照组。当分析结束，总结报告完成时，在临床试验总结会上再进行第二次揭盲，以明确 A、B 两组中哪一个为试验组，哪一个为对照组，即分组盲底。当双盲试验的设计不是 1:1 时，如试验组与对照组为 2:1 设计，只需要进行一次揭盲，或者第一次揭盲时分为 A、B、C 三组，第二次揭盲时宣布其中的两组为试验组，另一组为对照组。

紧急情况下，个别病例的揭盲规定：在临床试验方案中需明文规定在什么情况下，由什么人可以拆开应急信件，获知该编号的病人接受的是何种处理。一般，受试者若出现严重不良事件，或死亡，或需紧急抢救时，可由该中心负责研究的研究者报告监查机构及主要研究者是否需拆开应急信件。需要指出的是，一份应急信件的打开仅仅涉及一个病例的揭盲。

7. 双盲试验的终止和失效

一般来说，在双盲临床试验进行的全过程中，一旦全部的盲底泄盲，或者应急信件拆阅率超过 20% 时，就意味着双盲临床试验的失效，须重新进行一个新的临床试验。

8. 盲底生成和药物分装的报告文件

上述全部过程应由专人书写成文件形式，作为该临床试验的文件之一，保存起来。该文件也可书写成编盲和试验药物分装记录。其内容应包括申办者对药物的准备，药物的包装、用法、储存要求及药物发放办法，随机分组编码的产生，按每个受试者包装的药盒、应急信件，药品检验所对试验药与对照药的检验报告，盲底的保存及揭盲的规定和各个中心药盒分配的编号等。

4.4.2 双盲临床试验中须注意的问题和技巧

1. 编码的补充量

在形成临床试验的盲底后，不能随便拆阅、改动，因此在编码过程中需留有一定的补充量以备不时之需，主要通过两个方面进行。

(1)多编一些试验用药号,并备好其相应的试验用药和应急信件。一般,可多编20%的试验用药号,如样本量为200例时,可以随机产生1~240个编码,前200个编码是正式的用药编号,它是设计方案中关于各处理组病例的安排。当前200个用药编号发生意外情况时,如药物损坏、操作失误或入组病例因意外原因而被剔除时,可用后面的40个编码进行补充。

(2)在每个用药编号中多放置20%的药物用量。比如,每个受试者用药量为10mg/片,3次/日,1片/次,10天为1个疗程时,则除每个疗程所需实际用药数(30片)外多放置6片(20%)以应对意外情况。上述操作过程需要严格执行药物的管理与回收,在随访每个受试者时,需了解用药情况,及时收回多余的药物。当试验用药为针剂时,最好能收回已用过的安瓿以便确认病人的用药情况。

2. 双模拟技术

安慰剂对照常用于双盲试验,需要注意的是使用安慰剂的临床试验不一定是安慰剂对照试验,如在阳性药物对照试验中,有时为了保证双盲的执行,会采用双盲双模拟技巧,对试验药、阳性对照药同时制作安慰剂模拟剂,这样的临床试验是阳性药物对照试验,而不是安慰剂对照试验。在临床试验中,阳性对照药通常选用已知的标准药物。由于用上市药品作为阳性对照药时不允许改变其原有的制剂、外观,如果试验药与阳性对照药在色、香、味、形、包装等方面相同、不易分辨,则容易实行盲法,而如果两药在外观、剂型等方面不一致,很容易被区分,则难以进行盲法试验。此时,可采用双盲双模拟法加以解决。

如果试验药与阳性对照药剂型不同,如试验药为片剂,而对照药为胶囊,为了执行双盲,可使用双模拟技术,由申办者制备一个与试验药完全相同的片剂安慰剂,称为试验药的安慰剂;再制备一个与对照药外观相同的胶囊安慰剂,称为对照药的安慰剂。按编码结果,如果一个受试者属于试验组,则服用试验药加上对照药的安慰剂;如果属于对照组,则服用对照药加上试验药的安慰剂。各药和其安慰剂服用方法相同。从整个用药情况来看,每个入组病例所服用的药物、每日次数、每次用量都是一样的,由此保证了双盲的实施。

3. 胶囊技巧

当两种药物剂型相同,但对照药品是进口的异型片剂,申办者无法模拟制作安慰片剂时,可以采用胶囊技巧,即将试验药与对照药分别装入一个外形完全相同的胶囊中供受试者使用。实际操作时,每个受试者服用的是形状相同的胶囊,但胶囊内所包含的内容却分别是试验药与对照药。这里的主要问题是:①药片放入胶囊后可能会改变药代动力学的参数;②需告诫受试者、研究者或参与试验的医务人员不可以打开胶囊探视其内所包装的东西。

4. 基础治疗

当新药临床试验需要用安慰剂作为对照,但是从伦理、道德上看,对病人仅仅使用安慰剂而不给予有效治疗不妥时,可以考虑对所有的受试者采用一种标准治疗药物,又称基础治疗,该基础治疗具有一定的疗效和安全性。在此基础上,试验组加用新药,对照组加用安慰剂,但是其存在的问题是在增加了有效治疗之后,难以判断试验结果是新药的作用还是新药与基础治疗的联合作用。

除了上述问题之外,在双盲临床试验的实施过程中,还应当注意临床结果的及时收集与管理,避免因为各项指标上的明显差别而使得研究者或受试者推测出试验的分组情况,造成试验的破盲。

4.5 临床试验实例

某研究机构欲观察二甲双胍格列吡嗪片的降血糖的治疗效果和服用此药后的安全性,进行了一项临床试验。

4.5.1 研究设计

采用随机、双盲、双模拟、多中心、二甲双胍格列吡嗪片与二甲双胍片和格列吡嗪片联合应用平行对照的研究方法。

(1)多中心试验。本试验在4家医院同时展开，每个试验中心对30对(60例)符合纳入标准的受试对象进行该项临床试验。

(2)随机化分组。一个试验中心分配一个编号。对符合入组条件的每个患者按顺序分配初筛号。研究者在第二周(第一次随访)时发给患者相应编号药物。初筛号不能重复使用，受试者不能被重复筛选。

对初筛符合入选标准的患者(第一阶段)，观察2周。观察期末(第二阶段)，根据第二阶段入选标准，入选者按1:1比例随机分入试验组或对照组，进入12周的双盲期，于治疗后的第2、4、8、12周随访，以提供二甲双胍格列吡嗪片有效性和安全性评估的数据和资料，并由研究者提供下一随访期的药物。

(3)双盲。两组受试对象及药物有效性、安全性评估指标的观察人员均不知道病例服用哪种药物。

(4)阳性药物对照。出于伦理的考虑，本试验组采用二甲双胍格列吡嗪片，对照组采用二甲双胍片和格列吡嗪片的联合应用。

(5)平行试验。本试验的每个病例只接受一种药物的治疗，即或试验药物或对照药物，不能交叉使用，便于观察试验药物的使用。

4.5.2 样本含量估计

对新药有效性和安全性初步评价的Ⅱ期临床试验要求试验组一般不小于100例，本次试验结束时至少有100对，200例临床有效病例(其中100例患者服用二甲双胍格列吡嗪片，100例患者服用盐酸二甲双胍片和格列吡嗪片)，按照实际病例数的120%安排，共计120对(240例)。

4.5.3 病例选择

临床试验的入选病例必须满足一定的条件，即入选标准和排除标准，同时还应对病例出现何种情况需终止试验制定标准。

1. 入选标准

(1)第一阶段入选标准(0~2周)。

①了解试验全过程，自愿参加并签署知情同意书。

②Ⅱ型糖尿病患者(WHO标准，1999年)。

③年龄为30~70岁，男女不限。

④体重指数： $19\text{kg/m}^2 \leq \text{BMI} \leq 35\text{kg/m}^2$ 。

⑤入选前3个月内未参加任何药物试验。

⑥新诊断的Ⅱ型糖尿病，空腹血糖值为9.0~15.0mmol/L。

⑦既往已诊断的Ⅱ型糖尿病，饮食控制、运动治疗+磺酰脲类药物(单药最大剂量<6片/日)，或+二甲双胍，或+α-糖苷酶抑制剂，或+格列奈类药物(<6片/日)，或+噻唑烷二酮类药物，用药剂量稳定1个月以上，空腹血糖值为7.0~13.4mmol/L，或HbA1c>7.0%。

⑧入选前3个月内未使用胰岛素治疗。

⑨无严重心、肺、肝、肾疾病。

⑩育龄女患者有避孕措施。

⑪无其他系统性疾病(如结核、结缔组织病等)。

(2)第二阶段入选标准(0周)。

新诊断的Ⅱ型糖尿病人空腹血糖为 $8.0 \sim 14.0\text{mmol/L}$ 者入选;既往已诊断的Ⅱ型糖尿病人空腹血糖与初筛时差值 $\leq 2.0\text{mmol/L}$ 者入选;差值 $> 2.0\text{mmol/L}$ 者再观察1~2周,与上一次空腹血糖差值 $\leq 2.0\text{mmol/L}$ 者入选。

2. 排除标准

有以下情况之一不能纳入本研究。

(1)孕妇、准备妊娠或哺乳期妇女。

(2)不能合作者。

(3)肝功能异常:ALT、AST为正常上限2倍以上者。

(4)肾脏损害:血肌酐超过正常上限的90%者。

(5)严重慢性胃肠道疾病患者,如胃溃疡、十二指肠球部溃疡、慢性结肠炎等。

(6)严重心脏病患者,如心力衰竭、不稳定型心绞痛、心肌梗死者。

(7)血压 $\text{SBP} > 160\text{mmHg}$ 和/或 $\text{DBP} > 95\text{mmHg}$ 者。

(8)有明显血液系统疾病者,如中重度贫血、血小板减少性紫癜、血液系统恶性病变等。

(9)有其他内分泌系统和代谢性疾病,如甲亢、皮质醇增多症等。

(10)对试验药物过敏或频发低血糖者。

(11)有中重度糖尿病慢性并发症者,如视网膜病变Ⅱ期以上、尿蛋白 $> ++$ 、糖尿病足等。

(12)需胰岛素治疗糖尿病酮症酸中毒、高渗性非酮症昏迷等患者。

(13)有手术、严重外伤等应激情况者。

(14)有慢性缺氧性疾病如肺气肿、肺心病等。

(15)明显听力下降或耳鸣者。

(16)有精神疾患,药物或其他物品滥用(包括酗酒)者。

(17)过去12周内献血超过 500mL 或至少失血 500mL ,或在过去8周内有输血史者。

3. 退出标准

有以下情况之一者应退出本研究。

(1)使用试验药物时出现过敏反应或其他严重不良事件及不能耐受者。

(2)不能按要求进行本试验者。

(3)反复低血糖者(由主管医生根据方案决定)。

4. 终止试验标准

(1)因低血糖中止治疗的原则如下:

①不能解释的2级低血糖事件。

②不能解释的频繁(> 3 次/周)临床1级低血糖事件。

③在整个试验期内出现不能解释的,持续(> 3 次/周)无症状性低血糖,血糖 $\leq 2.7\text{mmol/L}$ (50mg/dl)或3次血糖结果 $< 3\text{mmol/L}$ 。

(2)随机后因高血糖中止试验的原则如下。

①高血糖症状恶化,如在不伴随疾病发生情况下出现尿频、烦渴、体重减轻或其他因高血糖恶化而出现的偶发情况。

②在医院试验室经两次测量确定的无症状的 $\text{FPG} > 13.3\text{mmol/L}$ (240mg/dl),无伴随疾病。

(3)妊娠：

如果女性患者在试验期间怀孕，必须停止试验用药，并给予患者适当的降糖治疗。

(4)在随机前中止试验：

患者在清洗期中止试验，需进行中止试验时所要求的全部检查，停用清洗期发放的药物，并记录伴随用药或治疗。

(5)在随机后中止试验：

对于在随机后的双盲治疗 12 周内中止试验的患者，需完成第 12 周随访的评估，停用试验药物，记录不良事件、合并用药或伴随治疗和低血糖评估。

4.5.4 给药方案

1. 第一阶段(清洗期)

第一次用药在 0 ~2 周随访后的第一餐的餐前进行；最后一次服药在 0 周随访的前一天的晚餐前服用。

早、晚餐前各服观察用药 1 袋。

2. 第二阶段(双盲治疗期)

第二阶段共 12 周。第一次用药在 0 周随访后的第一餐的餐前服用。最后一次服药在第 12 周随访前一天的晚餐前服用。

初始用药(0 周)：早、晚餐前各服访视用药 1 袋。

服药 2 周后：空腹血糖 3.5 ~8.3mmol/L 者，剂量不变；>8.3mmol/L 者，早、晚餐前各服访视用药 1 袋，午餐前服加量用药 1 袋；<3.5mmol/L 者，3 天内复查血糖，仍<3.5mmol/L 者中止试验。

服药 4 周后：空腹血糖 3.5 ~8.3mmol/L 者，剂量不变；>8.3mmol/L 者，早、晚餐前各服访视用药 1 袋和加量用药 1 袋；<3.5mmol/L 者减去一次加量药物，如果未服加量药物则 3 天后复查血糖，仍<3.5mmol/L 者中止试验。

服药 8 周后：空腹血糖 3.5 ~8.3mmol/L 者，剂量不变；>8.3mmol/L 者，早、晚餐前各服访视用药 1 袋和加量用药 1 袋，午餐前服加量用药 2 袋；<3.5mmol 者减去加量药物，如果未服加量药物则 3 天后复查血糖，仍<3.5mmol/L 者中止试验。

最大服药剂量：试验组为盐酸二甲双胍 1500mg、格列吡嗪 15mg/日，对照组中盐酸二甲双胍为 1500mg/日，格列吡嗪为 15mg/日。

用药过程中出现低血糖(≤2.7mmol/L)反应者，如正在服访视用药，则中止试验；如正在服加量药物，则减去一次加量药物；如加量药物均减去后仍有低血糖则中止试验；如因饮食、运动不当所致低血糖，应纠正诱因后观察，3 天后复查血糖，仍<3.5mmol/L 者中止试验。

在随访当日早上不要服药、不吃早餐，在收集完所有需空腹的试验室血样后再服用早上的药品。

4.5.5 疗效指标

1. 有效性评价

(1)主要指标：空腹血糖、餐后 2h 血糖及 HbA1c。与试验前相应指标比较；该血糖及 HbA1c 值与使用对照组所测值进行比较；评价是否与对照药有相同疗效。

(2)次要指标：两组血脂水平比较。

2. 安全性评价

(1) 试验室检查异常。

①0周、2周、4周、8周、12周进行体格检查。

②12周进行肝肾功能、血常规、尿常规及12导联心电图的检查。

(2) 不良事件发生率。

①试验药品的常见不良事件：低血糖、消化道症状。

②本次试验中不良事件的判定：本次试验中，自患者签署知情同意书并入选试验开始至最后一次随访期间，发生的任何不利医疗事件，无论与试验药物是否有因果关系，均判定为不良事件。

4.5.6 数据录入与核查

1. 数据录入与修改

数据录入与管理由专门数据管理员负责。数据管理员采用 Excel 软件，进行数据录入与管理。为保证数据的准确性，应由两个数据管理员独立进行双份录入并校对。

对病例报告表中存在的疑问，数据管理员将产生疑问解答表(DRQ)，并通过临床监查员向研究者发出询问，研究者应尽快解答并返回，数据管理员根据研究者的回答进行数据修改、确认与录入，必要时可以再次发出 DRQ。

2. 数据锁定

在盲态审核并确认建立的数据库正确后，由主要研究者、申办者、统计分析人员对数据进行锁定。对锁定后的数据文件不再做改动。对数据锁定之后发现的问题，经确认后在统计分析程序中进行修正。

3. 揭盲和数据处理

在研究数据全部录入并锁定后，由保存盲底人员做第一次揭盲，并将数据库交统计分析人员按统计计划书要求进行统计分析。完成统计分析后，由统计分析人员写出统计分析总、分报告，最后再由保存盲底人员进行第二次揭盲，交本试验的主要研究者写出研究报告。

4.5.7 统计分析

1. 数据集人群确定

按意向治疗原则和符合方案数据集分别确定数据集人群，进行意向性分析和符合方案分析。

(1) 意向性分析：对所有经随机化分组，用了试验药品，并至少有一次基线和基线后观察评价的全部病历进行意向性分析，意向性分析针对疗效和不良事件进行。在意向性分析中，对于在试验中途脱落的病例，其脱落后各时间观察点的指标数据缺失值，由相应指标的最后一次观察值结转；对于完成病例试验中间所缺失的数据，由最近一次非缺失的相应指标数值结转；对于经随机化入组后脱落的病例，且无用药后可评价的数据，对其缺失数据不做任何结转，按缺失数据处理。

(2) 符合方案分析：对所有进入随机化分组，符合试验方案，且依从性好，并按试验研究方案完成试验的病例。进行所有疗效指标的分析等。其缺失数据不做结转，均按缺失数据处理。

对主要疗效指标分析采用意向性分析和符合方案数据集分析，次要疗效指标分析采用意向性分析，基线资料分析和安全性分析采用意向性数据集。

2. 统计分析方法

统计分析采用国际通用 SAS(6.12) 统计分析软件进行计算。本试验研究的统计学检验均采用双侧检验，以 $P \leq 0.05$ 作为判断差别有统计学意义的标准。

(1)数据的描述性分析:对定量资料主要是描述均数、标准差、最大值和最小值;对定性资料(含等级资料)主要描述频数分布、构成比及百分率等。通过在计算机上对数据库试运行,按资料的类别对照临床试验方案,验证例数、指标数量、指标取值范围及数据的真实合理性。核实审查数据,看其有无逻辑差错。对于检出的错误,要查清错误来源,及时落实。

描述性分析对整体数据有一个基本的叙述,为进一步的统计分析提供信息和准备。对于计量资料进行正态性分布检验,以确定深入统计分析的方法。正态性检验时,检验水准 α 取0.05。正态分布的资料采用参数统计方法,非正态分布的资料采用非参数统计方法。

(2)数据资料比较的均衡性分析:在临床试验中,试验组与对照组除试验因素不同外,其他的影响因素在组间应均衡,这样组间产生的差异才是来自试验因素的作用。因此,首先分析组间的均衡性,即除试验因素以外因素的齐同性。组间均衡性的统计学检验,要求 $P>0.05$, P 值越大越好,因此计算出具体 P 值。一般对于正态性的计量资料用方差分析或 t 检验,对于非正态性的定量资料用非参数方差分析等。对于定性资料的计数和等级资料用卡方检验、精确概率法、Ridit分析或秩和检验等。

(3)基本指标的统计学分析:对于定量资料,若数据服从正态分布,选用参数检验法,如 t 检验、方差分析等。若数据非正态分布,则选用非参数检验法,如Wilcoxon秩和检验等。如需要通过数据变换,非正态性资料可转为正态性资料,可用参数检验方法。定量资料根据需要也可转变为定性资料。但定量资料转为定性资料会丧失部分信息,导致检验效能降低,因而不轻易将定量资料转为定性资料。对于定性资料的计数资料,一般对于双向无序 $R\times C$ 表资料,采用卡方检验,当表中理论频数小于5的格子数超过全部格子数的1/5时,用Fisher的精确检验;对于单向有序 $R\times C$ 表资料,采用秩和检验或Ridit分析。

(4)疗效结果指标分析:以上分析方法同样适用于疗效指标分析。对于正态性计量资料的组内自身比较采用配对 t 检验;对于非正态性计量资料的组内自身比较采用符号秩和检验等,以验证每组自身对照的疗效。

组间疗效比较应用终点数据和用药前后差值数据,应用 t 检验(正态性资料)或Wilcoxon秩和检验(非正态性资料)。

为避免中心因素对疗效的影响,应用Mantel-Haenszel分析,排除中心因素的影响,即使各中心同组(试验组或对照组),病人疗效不一致,其得出的统计结果仍为可靠。为消除疗效指标基线值的影响,应用协方差分析比较组间差异。

对于疗效评价的结果指标(一般是单向有序 $R\times C$ 表资料),应用秩和检验或Ridit分析。另外在组间的疗效比较分析中,要包括按试验方案实际完成病例分组(PP原则)的疗效比较和按随机化分组的疗效比较(ITT原则),即分析时不排除那些依从性差、不遵守试验安排和中途退出的病例。

(5)安全指标分析:不良事件发生率及试验室检验结果异常变化。不良事件按不良反应和无关不良事件分类。本研究从保守角度考虑,不良反应按不良事件与试验研究的药物“肯定有关”、“可能有关”及“肯定无关”划定。采用 χ^2 检验比较两组不良事件发生率,并列表描述本次试验所发生的不良事件;试验室检验结果在试验前后正常/异常的变化情况以及发生异常改变时与试验药物的关系。

依从性分析:依从性根据药品发还表记录的数据评价。

依从性按(80%、80%~120%、120%)分类,80%~120%为依从性良好。

4.6 诊断试验的评价

诊断试验(diagnostic test)是评价疾病诊断方法的临床试验。在临床医学中,医学诊断对患者的诊治过程及其预后具有重要作用。好的诊断试验无疑会产生好的临床效果。然而,在诊断某一病

种过程中往往会用到不止一种诊断试验，并且，随着医学科技的进步还将不断产生新的诊断试验，如何评价这些诊断试验，并从中选择较优者，除有赖于临床实践外，还需诊断试验的统计分析手段，才能使这一问题得到较好解决。

4.6.1 诊断试验的设计步骤

1. 确定金标准

所谓“金标准”是指临床医学界公认的可可靠的疾病诊断方法，也称标准诊断方法，它能正确地分有病和无病。临床上常用的金标准包括病理学检查(组织活检和尸体解剖)、外科学手术以及长期随访患者在临床上获得的肯定结论。通过待评价的诊断试验和金标准检测相同的受试对象，然后比较两种诊断方法的诊断结果，从而正确评价待评价诊断试验的准确性、可靠性等内容。

2. 盲法测试

首先，通过标准诊断的结果将试验对象分为实际患某病(病例组)与未患某病(无病组)两组，然后用待评价的诊断试验检测。测试过程要遵循盲法的原则，即观察者和受试者不能预先知道何为金标准确定的病例或对照，以免发生主观偏倚。

3. 试验资料的整理

研究对象经标准诊断和诊断试验测试后四种可能的情况：真阳性(实际患病者试验结果阳性)、假阳性(实际无病者试验结果阳性)、假阴性(实际患病者试验结果阴性)和真阴性(实际无病者试验结果阴性)。真阴性和真阳性的试验结果是正确的，假阴性和假阳性的试验结果是错误的。具体资料整理形式，见表 4-14。分别计算各种情况的发生概率，就可得出诊断试验真实性的指标。

表 4-14 诊断试验评价的 4 种可能的情况

诊断试验的结果	例 数		
	*：有病(D_+)	无病(D_-)	
阳性(T_+)	真阳性(a)	假阳性(b)	$a + b$
阴性(T_-)	假阴性(c)	真阴性(d)	$c + d$
合计	$a + c$	$b + d$	n

注：* 代表“标准诊断结果”。

4.6.2 诊断试验的评价指标

1. 真实性的评价

真实性，也称准确性，是指检测结果与真实情况符合的程度。

(1)灵敏度(sensitivity, S_e) 又称真阳性率，是实际患病且被诊断试验诊为患者的概率，即患者被判为阳性的概率，反映检出患者的能力，该值越大越好。

$$S_e = \frac{a}{a + c}$$

(4-15)

(2)特异度(specificity, S_p) 又称真阴性率，是实际未患病则被诊断试验诊为非患者的概率，即非患者被判为阴性的概率，反映鉴别非患者的能力，该值越大越好。

$$S_p = \frac{d}{b + d}$$

(4-16)

(3)误诊率(mistake diagnostic rate, α) 又称假阳性率，是实际未患病而被诊断试验判为患者的概率，即非患者中被判为阳性的概率，反映将非患者诊断错误的可能程度，该值越小越好。

$$\alpha = \frac{b}{b + d} \tag{4-17}$$

(4)漏诊率(omission diagnostic rate, β) 又称假阴性率, 是实际患病而被诊断试验判为非患者的概率, 即患者中被判为阴性的概率, 反映将患者诊断错误的可能程度, 该值越小越好。

$$\beta = \frac{c}{a + c} \tag{4-18}$$

(5)符合率(π)分为粗符合率(crude agreement)和调整符合率(adjusted agreement), 说明诊断检试验阳性与阴性结果均正确的百分比, 反映正确诊断患者与非患者的能力。

粗符合率

$$\pi = \frac{a + d}{a + b + c + d}$$

调整符合率

$$\pi = \frac{1}{4} \left(\frac{a}{a + b} + \frac{a}{a + c} + \frac{d}{b + d} + \frac{d}{c + d} \right) \tag{4-19}$$

(6)Youden 指数(Youden’s index, YI)将误诊率和漏诊率综合起来, 定义 YI 为

$$YI = 1 - (\alpha + \beta) = 1 - \alpha - \beta \tag{4-20}$$

其值在 0 ~ 1 之间变动, 其值越大, 诊断试验的真实性越好。

(7)比值比(odds ratio, OR)将敏感度与特异度综合起来, 其值越大, 诊断值越高。但其适用于样本含量较大且四格表内的数均不为零的情况。

$$OR = \frac{S_e}{1 - S_e} \times \frac{S_p}{1 - S_p} = \frac{ad}{bc} \tag{4-21}$$

(8)似然比(likelihood ratio, LR)。似然比为病人中出现某种检测结果的概率与非病人中出现相应结果的概率之比, 有阳性似然比和阴性似然比。阳性似然比(positive likelihood ratio, PLR)是指真阳性率与假阳性率之比, 说明病人中出现某种检测结果阳性的概率是非病人的多少倍, 即一项试验按某已定标准判断某病人结果为阳性。阴性似然比(negative likelihood ratio, NLR)是指假阴性率与真阴性率之比, 说明病人中出现某种检测结果阴性的概率是非病人的多少倍。

$$PLR = \frac{a/(a + c)}{b/(b + d)} \tag{4-22}$$

$$NLR = \frac{c/(a + c)}{d/(b + d)} \tag{4-15}$$

2. 可靠性的评价

可靠性, 也称重复性, 指在相同条件下, 同一观察者(或不同观察者)用同一种检测方法重复检测同一批或者同一位受试者, 各次结果之间的一致程度。定量和定性资料可分别通过计算组内相关系数(intra-class correlation coefficients, ICC)、kappa 系数进行评价。ICC 值和 kappa 系数均介于“0 ~ 1”之间, ICC/kappa > 0.75, 一致性“佳”, ICC/kappa < 0.4, 一致性“差”, 介于“0.4 ~ 0.75”之间, 则认为一致性良好。ICC 的计算可通过随机区组设计的方差分析来完成, 公式如下:

$$ICC = \frac{MS_b - MS_e}{MS_b + (k - 1)MS_e + k(MS_k - MS_e)/n} \tag{4-23}$$

式中, MS_b 为区组间(即观察对象)均方, MS_e 为误差均方, MS_k 为处理组间均方, k 为处理组数, n 为样本含量。

3. 收益的评价

(1)预测值(predictive value, PV), 指诊断试验结果与实际(标准诊断结果)符合的概率。上述几项指标是在已知受试者状态(患病或未患病)的条件下估计诊断正确或错误的可能性有多大。在另一种情况下, 我们事先并不知道受试者处于何种状态, 但希望依据诊断结果判断受试者处于某一

种状态的可能性有多大。诊断结果阳性者的患病概率称为阳性预测值(positive predictive value, PV_+)。诊断结果阴性者未患病的概率称为阴性预测值(negative predictive value, PV_-)。它们的计算公式为

$$PV_+ = \frac{a}{a + b} \tag{4-24}$$

$$PV_- = \frac{d}{c + d} \tag{4-25}$$

【例 4-23】 用前列腺酸性磷酸酯酶放免(RIA-PAP)诊断前列腺癌。对 113 名前列腺癌患者和 217 名非患者的诊断试验结果见表 4-15，计算评价诊断试验的各项指标。

表 4-15 RIA-PAP 对前列腺癌的诊断试验的结果

RIA-PAP 结果	患 者 数	非 患 者 数	合 计 数
+	79	13	92
-	34	204	238
合计	113	217	330

【解】

灵敏度 $S_e = \frac{79}{113} = 69.91\%$

特异度 $S_p = \frac{204}{217} = 94.01\%$

误诊率 $\alpha = \frac{13}{217} = 5.99\%$

漏诊率 $\beta = \frac{34}{113} = 30.09\%$

粗符合率 $\pi = \frac{a + d}{a + b + c + d} = \frac{79 + 204}{330} = 85.76\%$

调整符合率 $\pi = \frac{1}{4} \left(\frac{a}{a + b} + \frac{a}{a + c} + \frac{d}{b + d} + \frac{d}{c + d} \right) = 83.88\%$

Youden 指数 $YI = 1 - \alpha - \beta = 1 - \frac{13}{217} - \frac{34}{113} = 63.92\%$

比值比 $OP = \frac{ad}{bc} = 36.46$

阳性似然比 $PLR = \frac{a/(a + c)}{b/(b + d)} = 6.01$

阴性似然比 $NLR = \frac{c/(a + c)}{d/(b + d)} = 0.16$

阳性预测率 $PV_+ = \frac{79}{92} = 85.87\%$

阴性预测值 $PV_- = \frac{204}{238} = 85.71\%$

(2) 效果。

可从生物学和社会经济学效果来考察。生物学效果：生存率是评价用于人群筛检的诊断试验效果比较理想的指标。社会经济学效果：可从成本效果、成本效益、成本效用三方面进行评价，其中最重要的是成本效益分析(详见关于卫生经济学的书籍)。

4.6.3 联合试验

联合试验是指采用多个筛检试验检测一种疾病，达到提高筛检试验灵敏度或特异度的目的，以满足提高筛检试验真实性的需要，有串联试验和并联试验两种。串联试验(serial test)，也称系列试验，是指采用几种筛检方法检测疾病，只有全部检测均为阳性者才判为阳性，凡有一项检测结果为阴性即判为阴性。并联试验(parallel test)也称平行试验，是指采用几种筛检方法检测疾病，凡有一项检测为阳性者即判为阳性，所有检测均为阴性才判为阴性。联合试验结果判定，见表 4-16。

表 4-16 联合试验结果判定方法

试 验 方 法		串联试验结果		并联试验结果
<i>A</i>	<i>B</i>			
+	-	-		+
-	+	-		+
+	+	+		+
-	-	-		-

如果两个筛检试验彼此完全独立，则可以采用下列公式计算联合试验的灵敏度和特异度。

并联试验的灵敏度 = *A* 灵敏度 + (1 - *A* 灵敏度) × *B* 灵敏度

并联试验的特异度 = *A* 特异度 × *B* 特异度

串联试验的灵敏度 = *A* 灵敏度 × *B* 灵敏度

串联试验的特异度 = *A* 特异度 + (1 - *A* 特异度) × *B* 特异度

由联合试验灵敏度和特异度的计算公式可见，串联试验提高特异度，并联试验提高灵敏度。串联试验和并联试验实例，见表 4-17，通过尿糖和血糖联合筛检糖尿病病人。联合试验时，应先进行特异度高的试验，后用灵敏度高的试验，这样可使筛检的效率更高。

表 4-17 单项试验及联合试验筛检糖尿病结果

筛 检 试 验	试 验 结 果	糖 尿 病 病 人	非 病 人	灵 敏 度 (%)	特 异 度 (%)	假 阴 性 率 (%)	假 阳 性 率 (%)
单项尿糖	阳性	131	31	65.83	99.59	34.17	0.41
	阴性	68	7610	—	—	—	—
单项血糖	阳性	150	32	75.38	99.58	24.62	0.42
	阴性	49	7609	—	—	—	—
串联试验	阳性	117	21	58.79	99.73	41.21	0.27
	阴性	82	7620	—	—	—	—
并联试验	阳性	164	42	82.41	99.45	17.59	0.55
	阴性	35	7599				

4.6.4 受试者工作特征曲线

1. 受试者工作特征曲线

受试者工作特征曲线(receiver operator characteristic curve, ROC 曲线)是以试验灵敏度(也叫真阳性率,记为 TPR)为 *y* 轴,以(1 - 特异度)(也叫假阳性率,记为 FPR)为 *x* 轴,由诊断试验不同临界点下的灵敏度和假阳性率产生图中各个点,采用线段连接而成的线图。如金标准测定的 360 例受试对象,其中 230 例心肌梗死者,58 例无心肌梗死,某心内科医生又通过受试对象的肌酸磷酸激酶浓度(CPK)进行诊断,结果见表 4-18。将诊断临界点分别确定在 ≥280、≥80、≥40 和 ≥1 时,分别计算各临界点下的真阳性率(TPR)和假阳性率(FPR),然后得到该 CT 影像诊断试验的 ROC 曲线,见图 4-1。

表 4-18 360 例受试对象在不同 CPK 水平下的分布

金 标 准	例 数				合 计
	CPK 水平: ≥280	80 ~279	40 ~79	1 ~39	
有心肌梗死	97	118	13	2	230
无心肌梗死	1	15	26	88	130

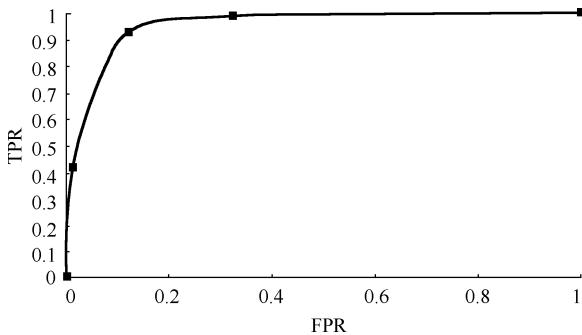


图 4-1 360 例受试对象 CPR 诊断试验的 ROC 曲线

绘制 ROC 曲线的 SAS 程序如下：

```
data nkkysj6_4_1;
input TPR FPR@@;
cards;
0 0 0.01 0.42 0.12 0.93 0.32 0.99
;
run;
Proc gplot;
Plot FPR* TPR;
Run;
```

程序说明：数据步输入不同临界点下的灵敏度和特异度，过程步绘制 FPR 为横坐标轴、TPR 为纵坐标轴的 ROC 曲线。

2. ROC 曲线的作用

- (1) ROC 曲线能很容易地查出诊断试验在任意界值时的对疾病的识别能力。
- (2) 选择最佳的诊断界限值。ROC 曲线越靠近左上角,试验的准确性就越高。最靠近左上角的 ROC 曲线的点是错误最少的最好阈值，其假阳性和假阴性的总数最少。
- (3) ROC 曲线下面积可以反映某个诊断试验的价值大小。AUC ≤ 0.7 ,表示诊断价值较低；0.7 < AUC ≤ 0.9 ,表示诊断价值中等；AUC > 0.9 ,表示诊断价值较高。
- (4) 两种或两种以上不同诊断试验对疾病识别能力的比较。在对同一种疾病的两种或两种以上诊断方法进行比较时，可将各试验的 ROC 曲线绘制到同一坐标中，以直观地鉴别优劣，靠近左上角的 ROC 曲线所代表的受试者工作最准确。也可通过分别计算各个试验的 ROC 曲线下的面积 (AUC) 进行比较，哪一种试验的 AUC 最大，则哪一种试验的诊断价值最佳。

4.6.5 诊断试验临界点的确立方法

对同一种疾病，应用不同的临界点进行诊断会得到不同的结果。因此，要结合临床实际选择符合专业要求、假阳性率(误诊率)和假阴性率(漏诊率)都较小的临界点。确定临界点常用的方法：正态分布法、百分位数法、ROC 曲线法、临床确定法。

1. 正态分布法

目前临床上大多采用“均数 ± 2 个标准差”作为临界点，在范围之内为正常，范围之外为异常。采用此方法确定临界点的诊断试验数据要求服从正态分布。

2. 百分位数法

由于多数诊断试验测定的频数不服从正态分布，故多数学者主张采用百分位数法，制定双侧临

界点时，计算第 2.5 百分位数和第 97.5 百分位数，在两个百分位数的范围之内为正常，范围之外为异常。

3. ROC 曲线法

根据诊断试验绘制的 ROC 曲线，找到灵敏度和特异度都最大点，也就是曲线上最靠近左上方的点，作为最佳临界点，然后根据临界值判定正常和异常。

4. 临床确定法

在临床上，一个诊断试验的测定值达到什么水平才需要治疗，常根据人群调查中是否系危险因素来判断。例如，血清胆固醇水平超过 6.5mmol/L 水平时，发生冠心病的危险性明显升高，而低于此水平危险就没有统计学意义，故将胆固醇水平定位超过 6.5mmol/L 水平为异常。

4.6.6 诊断试验评价中常见的偏倚

1. 工作偏倚

对试验出现阳性结果的病人才决定进一步采用金标准加以确诊，而对阴性结果的病人不再做进一步检查，结果造成缺乏假阴性的资料。

2. 缺乏非病人群试验结果的信息造成的偏倚

例如有腰背痛的病人做磁共振检查，发现许多病人椎间盘膨出，故常用此结果来解释腰背痛的原因，并给予治疗。另一篇文章在 98 例无腰背痛的人中做磁共振检查，结果发现有 2/3 无症状者也有椎间盘膨出，其发生率仅略低于有症状者，两者统计学上无差别。

3. 由不明确结果者引起的偏倚

诊断试验常有结果不明确者或呈中间结果者，在分析资料时，作者常将它们剔除，从而造成结论不真实。

4. 试验室测量偏倚

诊断试验操作不正规，缺乏质量控制，没有进行重复性测定，没有对观察者、仪器在不同时间测定的变化加以分析等原因造成的结论不真实。

5. 参考试验偏倚

参考试验偏倚 (references test bias) 是指选择金标准不妥当造成的偏倚。例如评价 B 超对胆石症的诊断价值，采用口服胆囊造影作为诊断胆石症的金标准，实际作为金标准的口服胆囊造影要比 B 超结果差，从而造成灵敏度、特异度评价结果的不正确。

(任艳峰)

第5章 内科科研调查设计与数据分析

5.1 调查设计的基本内容

在研究工作中,若研究者在没有采取任何干预措施条件下,对研究对象进行观察和记录结果,这种研究方法称为调查(survey)。调查研究被广泛用于医学、社会学、教育学等领域,人口普查、生活状况调查、市场调查、民意测验等均是调查研究的具体应用。随着疾病谱的改变和现代医学模式从“生物—医学”模式向“生物—心理—社会”医学模式的转变,医学调查研究的范围也在相应地扩大,从急性病到慢性病,从传染病到非传染病,从疾病到健康,从疾病控制到卫生政策的制定等各个环节。

内科学作为医学的一个重要领域,调查研究也广泛地应用于内科学,常应用于疾病认知(如对糖尿病的认知情况)、患病特点(如高血压的患病现状)、病因学探讨(如心脏病的病因调查)、疾病预防对策(如支气管炎的预防)等各领域。

调查研究按调查时间的顺序一般可分为以下三类:横断面调查研究、追踪性调查研究、回顾性调查研究。

调查设计(survey design)是对调查研究所做的周密计划,它包括调查研究资料的收集、整理和分析全过程的统计设想和科学的合理安排,所以做好调查设计是内科卫生工作者的一项基本功。

5.1.1 调查的概念及其特点

调查(survey)是指在没有任何干预措施的条件下客观地观察和记录研究对象的现状及其相关特征。

在调查中,欲研究的对象及其相关特征(包括研究因素和非研究因素)是客观存在的,不能采用随机分配的方法来平衡或消除非研究因素对研究结果的影响,这是调查研究区别于试验研究的最重要特征。当然,对非研究因素的控制可以在调查分析阶段通过标准化法、分层分析以及多因素统计分析等方法得以实现,而不是在调查阶段。

5.1.2 调查设计的基本原则与内容

1. 明确调查目的

每一项调查,必须有明确的调查目的;调查目的一般可分为调查的总目的和具体目的;调查目的是选定调查指标的依据。

2. 确定调查对象和观察单位

调查对象是指根据调查目的确定的所要研究的“同质总体”。观察单位是指构成调查对象的各个单位或各个“个体”,观察单位可以是一个人、一个家庭或一个群体。例如,某地区为可疑食管癌高发区,拟进行现场调查,目的是摸清该地区食管癌的病情现状及具体地理分布,为病因研究提供线索,并为防治工作提供依据。在该调查研究中,调查对象为该地区某年的全部常住人口,观察单位是该地区的每个人。

3. 确定调查方法

应根据研究问题的性质、客观条件、研究目的等选择合适的调查方法。按调查的涉及面，一般可分为普查(overall survey)和抽样调查(sampling survey)。普查也称全面调查(complete survey)，是对调查范围内的全部观察对象(总体)进行调查，一般用于了解总体在某一特定“时点”的情况。抽样调查是一种非全面调查，是从总体中抽取一定数量的观察单位组成样本，然后根据样本信息来推断总体特征。抽样调查是医学科研中最为常用的方法。调查方法还可按调查的内容发生的时间，分为横断面调查(cross-sectional study)和纵向调查(longitudinal study)；按资料的来源，可分为现场调查和利用现有资料两种；按调查方式，可分为面对面调查和非面对面调查(信访、电话采访等)两种。

4. 确定调查指标和变量

调查目的是选定调查指标的依据，调查指标是调查目的的具体体现。设计时，应将调查目的转化为具体的调查指标。调查指标可分为客观指标和主观指标，还可分为定性指标和定量指标。指标的设立应注意灵敏性、特异性和客观性，并紧扣研究目的，做到少而精。

5. 调查工具和调查表

(1)调查工具。调查工具(instruments)可分为两类：一类是“硬”工具，另一类是“软”工具。例如，尺、秤、温度计等是“硬”工具；调查表和问卷等是“软”工具。调查工具必须标准化，要防止系统误差。工具的使用和调查结果的记录也必须标准化。调查表和问卷是调查过程中最常用的调查工具。

(2)调查项目。根据调查指标确定对每个观察单价的调查项目(item)，包括分析项目和备查项目。

①分析项目：是根据调查目的确定的直接用于整理和数据分析的指标以及分析时排除混杂因素影响所必需的内容。例如整理和分析食管癌调查指标，就必须调查食管癌死者的诊断结果、性别、死亡实足年龄(岁)、诊断食管癌的类型、身高、体重、是否喜欢吃滚热的东西、平时是否喜欢吃咸菜等几个项目，这里的性别、年龄项目还可用于计算食管癌调整病死率。

②备查项目：是为了保证分析项目填写得完整和正确，便于对其核查、补填和更正而设置的，通常不直接用于分析，如列出被调查者的姓名、性别、住址、电话以及调查者和调查日期等。列出死者的姓名和地址，有助于确定观察单位和核查；列出死者的出生年、月，可结合死亡年、月核算死时实足年龄；列出调查者和调查日期，有助于查询调查情况和明确责任。

调查项目要精简，分析项目一个也不可少，备查项目也不宜多，不必要的项目坚决不要。项目的定义要明确，提法要通俗易懂，使人不致误解。如疾病分型、正常或异常的界限，都应明确规定，不要模棱两可。

(3)调查表和问卷的设计。调查表和问卷(questionnaire)结构一般可划分为四个部分，分别为“说明部分”、“填写说明”、“核查项目”、“调查项目”。“说明部分”主要说明调查目的，以取得调查对象的合作；“填写说明”是为了保证所有调查员和调查对象均能对调查项目及填写方法正确理解、统一认识而编写的；“核查项目”这一部分是调查目的无关、不向调查对象询问的质量控制项目，如调查员姓名、调查日期、复核结果、未调查原因等；“调查项目”部分是由调查对象填写的部分，是调查的核心内容。

(4)问题的形式。根据问题答案的形式，问题可分为开放型和封闭型两类。开放型问题对问题答案不加任何限制，由调查对象对问题自由回答，适于调查者不清楚答案如何以及答案很多的情况，或事先不能确定回答的范围以及预调查。封闭型问题是根据问题可能的答案，提出两个或多个固定答案供调查对象选填，常用“是与否”或多项选择的形式。封闭型问题只能得到分类资料或等

级资料,而开放型问题有时可得到数值变量资料。问题答案形式可根据具体情况加以选择。

(5)问题设计应注意的问题。应尽量避免术语,避免含糊的提问用词,避免双重问题,避免诱导或强制,敏感问题的调查要有专门技巧。

具体调查表的设计和评价相关内容请见后面章节。

6. 确定样本含量

确定样本含量的意义如下:

- (1)可以控制统计量的抽样误差,样本含量越大,标准误越小。
- (2)提高估计的精度,增大样本含量是控制可信区间的宽度的有效办法。
- (3)增大样本含量是控制统计分析中Ⅱ型错误的概率的有效措施。
- (4)表示抽样误差的指标(如标准误)的抽样误差也与样本含量有关。

在现场调查中,最常用的是估计总体均数及估计总体率时要求的样本含量。

估计总体均数的样本含量的计算公式为 $n = \left(\frac{t_{\alpha/2} \sigma}{\delta} \right)^2$, 式中 δ 为允许误差。 σ 为估计的标准差,

一般是从以前的研究资料中获得的。在算得 n 之前,自由度 df 不能确定, $t_{0.05/2}$ 仍是未知的,解决的办法是先以 $u_{0.05/2}$ 代替 $t_{0.05/2}$, 用迭代法求得 n 。

估计总体率的样本含量的计算公式为 $n = \frac{u_{\alpha/2}^2 p(1-p)}{\delta^2}$, 式中 δ 为允许误差;估计的 p 是一个范围,应该取其中最靠近 50% 的值;假定估计的 p 在 10% ~ 30% 之间,则取 $p = 0.30$;假定估计的 p 在 40% ~ 80% 之间,则取 $p = 0.50$;如果对 p 一无所知,则取 $p = 0.50$ 。

其他类型的调查研究样本含量的计算见后面章节。

7. 确定抽样方法

总体而言,抽样方法可以分为概率抽样和非概率抽样两大类。

(1) 概率抽样。

概率抽样(probability sampling),就是在抽样中必须使该总体中的每一个个体都有已知的或可计算的和非零的概率被抽中。常用的概率抽样方法包括简单随机抽样、系统抽样、分层抽样和整群抽样。各种抽样方法的抽样误差一般是:整群抽样 \geq 简单随机抽样 \geq 系统抽样 \geq 分层抽样。在应用大多数概率抽样方法时,确切的抽样框架非常重要。抽样框架(sampling frame),简单地说就是一份完整的可以用来抽样的名单。如果没有抽样框架,也就是说目标人群(总体)不明确,那么所得的调查结论很难说适用于什么人群。具体四种常用概率抽样方法的相关问题见后面章节的详细叙述。

(2) 非概率抽样。

非概率抽样(non-probability sampling),是指各个体被抽样抽中的概率是未知的和无法计算的。然而,一些非概率抽样方法,尽管不能按常规的理论来计算抽样误差和推断总体,但在特定条件下还是有用的。但在应用中,不能忘了它们的局限性,特别要注意结论的合适性。非概率抽样包括配额抽样、滚雪球抽样等抽样方法。具体常用非概率抽样方法的相关问题见后面章节的详细叙述。

8. 资料收集方式

资料收集式主要有三种,即直接观察法、访谈法和问卷法,有时可以结合使用。

(1)直接观察法:是调查人员到现场对调查对象进行直接观察、检查、测量或计数来取得资料。例如,某地区发生急性食物中毒,医务人员到现场进行体检、收集标本。直接观察法取得的资料比较真实可靠,但所需人力、财力较多。一般来说,对于客观指标的测量、临床检查等,可采取直接观察法,如儿童身高、体重的测量。

(2) 访谈法：是根据被调查对象的回答来收集资料，常用的访谈的方式有以下三种。

①通过调查者向被调查者单独面对面的口头询问，将答案填入调查表，此种方法称为“访问调查”。此法可以使被调查者对问题的理解与设计要求一致，保证调查资料的准确性，一般应答率较高。

②通常是采用邮件或者电话等方式对调查对象进行间接调查，如将调查表邮寄给被调查者，请他们填好后寄回，该方法节省人力、财力，但由于调查人与被调查者不见面，被调查者对调查问题的理解常发生与设计要求不一致的情况，失访较多，以致影响调查资料的质量。

③与多个被调查者进行访谈，即以召开知情人座谈会的方式来收集资料，目前常用的是小组座谈会或专题小组讨论(也称焦点组访谈)，它可以收集定量的和定性的资料，运用的是非标准化形式讨论和观察的办法；节省时间、金钱并且信息准确可靠。

(3) 问卷法：是调查者运用事先设计好的问卷向被调查者了解情况或征询意见，这是一种书面调查方法。所谓调查问卷，是研究者根据一定的研究目的，将调查项目转换成一系列的问题，再印制成统一的调查表形式，即调查问卷，由被调查者本人直接填写调查问卷。这种设计的主要优点是调查成本较低且保密性好；缺点是调查对象对问题的理解可能与设计要求不一致，以致影响调查的质量。

9. 调查员选择和培训

调查员应该经过选择和培训，培训分为理论培训和实践培训。调查员的工作量要合理，对调查员应有监督机制和质量控制措施。

10. 调查的组织与实施

现场调查与实验室研究有很大不同，现场条件有时十分复杂，而且情况多变，所以要求调查者考虑周密。调查的组织计划是调查研究得以顺利实施的重要保证。

(1) 预调查。

在正式调查之前，要在小范围内做预调查，预调查的方式(包括调查对象、调查方法、采访方式等)应与正式调查的方式相同。预调查的目的在于检验调查的实用性，并通过预调查反馈的信息做必要的修改。只有当预调查成功后，才能批量印制调查表并正式使用。

(2) 组织计划。

调查的组织计划包括组织领导、宣传动员、时间进度、调查员培训、任务分工与联系、经费运算、调查表和宣传资料的准备以及质量控制方案等。调查设计中必须对上述工作做出周密的计划安排，并认真执行。在实施现场调查时，尤其应注意原始资料的完整性和准确性，发现问题及时补查或修正。

11. 伦理道德问题

伦理道德问题可以来自某些调查的问题本身，也可以来自为获得有效、可靠的资料所用的方法。调查时，要注意知情同意(informed consent)和隐私的保护。知情同意是指在研究对象暴露于某种危险之中或丧失某种个人权益时，要征得研究对象同意。

5.1.3 标准化率

在对调查资料进行对比分析时，要注意组间的可比性。当两组(或多组)资料的内部各小组的率明显不同，而且各小组观察单位的构成比明显不同时，则不能直接比较两组的总率。这时，可采用一个“统一的标准”将两组(或多组)资料的内部构成比例调整一致后，分别计算出调整后的总率再做比较，这种方法称为率的标准化(standardization)。率的标准化有以下两种方法。

1. 直接标准化

直接标准化(direct standardization)是以有代表性的、人群数量大的组作为标准人群,用标准人群各小组观察单位数分别乘以被标化人群的各小组的阳性率(如发病率),得到被标化人群的理论阳性数。然后,理论阳性数除以标准人群总人口数,得到被标化人群的标准化阳性率。

2. 间接标准化

间接标准化(indirect standardization)是以标准人群各小组阳性率乘以被标化人群的各小组观察单位数,得到被标化人群的理论阳性人数。被标化人群的实际阳性人数除以理论阳性总人数,得标准化阳性率比值(如标准化发病率比值 standardized incidence ratio, SIR 或标准化死亡率比值 standardized mortality ratio, SMR)。SIR(SMR)乘以标准人群实际阳性率,得到被标化人群的间接标准化阳性率。

3. 计算符号及公式

率的标准化计算数据符号如表 5-1 所示。

表 5-1 率的标准化计算数据符号

组 别	被 标 化 组			标 准 组		
	观察单位数	阳 性 数	率	观察单位数	阳 性 数	率
1	n_1	r_1	p_1	N_1	R_1	P_1
2	n_2	r_2	p_2	N_2	R_2	P_2
3	n_3	r_3	p_3	N_3	R_3	P_3
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
i	n_i	r_i	p_i	N_i	R_i	P_i
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
k	n_k	r_k	p_k	N_k	R_k	P_k
合计	n	r	p	N	R	P

直接法计算公式: $p' = \frac{\sum N_i p_i}{N}$ 。

间接法计算公式: $p' = P \frac{r}{\sum n_i P_i}$ 。其中, $SMR = \frac{r}{\sum n_i P_i}$ 。

5.1.4 调查表和问卷的设计与评价

1. 调查表和问卷的设计

常用的调查工具包括调查表、问卷(questionnaire)和量表(scale)等。调查表调查的内容可以十分宽泛,这些内容可能是完全独立互不相关的,可用于评价研究人群的不同特征,如询问调查对象的吸烟史、生育史、体育锻炼情况和饮食嗜好等。问卷常用于测量人们的心理、态度和行为等特征,而量表是一种特殊的问卷,常用于描述研究对象某一方面的特征,虽然量表用多个条目(问题)从各个方面来描述该特征,但各条目一般都是相关联的。这里简述问卷设计的主要原则。

1. 问卷设计的原则

(1)适宜性原则。设计问卷时要考虑研究的需要,同时也要考虑被调查者的实际情况。所以,问卷设计的首要原则就是要从被调查看的角度出发,尽可能减少他们在填答问卷时的困难与麻烦,减少他们填答问题所需要的时间和精力。

(2)有效性原则。问卷的问题必须围绕研究目的和研究假设进行设计。凡是与研究目的及其理论假设无关的问题,都必须删去;可有可无的问题,一般不列入问卷;不知如何分析的问题,也不要提出。设计者对问卷的设计要有一个总体框架.对设计的每一个问题所起的作用要十分清楚;对一个理论假设需要哪些指标来测量,也应十分明确。

(3)可行性原则。问卷调查需要被调查者的密切合作;因此,在设计问卷时,问卷中的问题必须符合被调查者回答问题的能力和意愿。问卷的问题要简洁,语言要通俗易懂,使被调查者能够顺利地完成任务。同时应考虑完成调查的时间要合适,问卷内容太多、调查时间太长,都可能影响调查质量。

2. 问卷的构成

(1)问卷的标题。问卷的标题是概括说明调查的研究主题。标题应简明扼要,易于引起被访者的兴趣。例如“居民营养与健康状况调查个人健康情况调查表”、“菜市居民肿瘤知晓认知与相关行为调查表”、“某地农村妇女生殖道感染调查表”、“癌症病人生活功能量表”等,不要简单采用“问卷调查”这样模糊的标题,它容易引起被访者因不必要的怀疑而拒绝回答。也不要将研究目的直接表达出来,如“饮酒与高血压关系调查表”,这样容易造成被访者的问答偏倚。

(2)问卷说明部分。问卷说明常常以简短的指导语或说明信的形式出现,旨在向被调查者说明调查的目的和意义。问卷说明可包括以下内容:

①说明调查者的身份,争取被调查者的信任和合作。说明中可以留下研究单位的联系方式,供被访者咨询和提出意见时使用。

②让被调查者了解调查的目的和意义,作用是激励研究对象的责任感,以负责的态度参与调查。

③请求被调查者合作。

④保密说明:为了使被访者消除顾虑,说明解释调查结果仅用于群体的科学研究,个人的隐私绝对保密,保证结果不用于其他用途等。

对于较大规模的调查,问卷项目较多,自填式问卷还必须有填表说明及其他事项说明等。填表说明的作用是解释问卷中某些项目的含义,指导被访者或调查者如何填写,例如对选择答案的填写方式和开放式答案的填写方式的说明等。问卷说明一般放在问卷开头,通过它可以使被调查者了解调查目的,消除顾虑,并按一定的要求填写问卷。

(3)问卷正文部分。正文部分是调查问卷的核心部分,主要包括被调查者信息、调查项目、核查项目三部分。

①被调查者信息。是指被调查者基本社会人口学特征,如性别、年龄、民族、婚姻状况、文化程度、职业等。这些项目一般用于对调查资料进行分组分析,从而探讨这些因素对调查主要结果的影响。

②调查项目。调查项目是调查问卷的核心内容,是本研究中需要调查的一些关键指标,是研究者需要分析的主要内容,也是问卷中最重要的部分,它主要以提问的形式让被访者回答。应尽量采用客观性强、灵敏度高、准确性好的定量指标。设置的项目要精选、定义要明确、问题要简明易懂、尽量做到不加说明或者稍加说明也能达到统一的标准。

③核查项目。是为了保证分析项目的完整性和正确性,设置便于核查和更正的项目。例如被调查者编号、姓名、家庭地址、工作单位、电话等,这些项目对分析可能不重要,但对资料的核查和对象的随访是重要的。

(4)编码。包括问卷编号、调查项目编号和问答选项编号。对于正规的问卷,还有记录框,将要录入计算机的各种数据和编码填于其中,以便于录入计算机。

(5)作业记载。在问卷的最后常需附上调查员的姓名、访问日期、时间等,以明确调查人员完成任务的情况,有助于查询和明确责任,减少调查问卷的填写缺失率和无效问卷数,保证调查质量。

3. 问卷的类型

根据设计的类型,调查问卷可以分为三个类型:开放型问卷、封闭型问卷和综合型问卷。

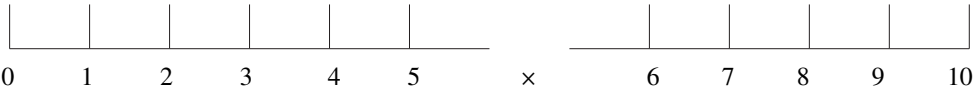
(1)开放型问卷:围绕调查研究目的提出一些问题,让调查对象自主说出自己的情况和想法,无备选答案,每个项目之后需留足必要的空白;被调查者根据自己的理解自由回答。

例如:填写体重_____kg、身高_____cm;您以往有何既往病史_____。
您对青少年吸烟有什么看法?_____
您对预防青少年近视有何建议?_____

开放型问卷优点是适于设计者不了解答案有哪些或答案难以分别列出的情况,缺点是获得的分类资料不便汇总。

(2)封闭型问卷:针对某一项目提供备选答案,供调查对象根据自己的情况在规定的范围内选择填写,不能自由发挥;是最常用的一种问卷形式,适合大范围的调查和研究。优点是答案标准化,容易回答,节省时间,拒答率低,便于对资料的统计汇总和分析。缺点是设计过程复杂,填答方式受限,事先规定的备选答案对被调查者的创造性有所限制,不利于发现新问题。当被调查者不理解问题或觉得没有合适备选答案时候,可能盲目或随便填写,造成调查结果偏离真实情况。答案的设置有以下三种情况:

- ①无序定性回答:列出所有可能的答案,供调查对象选择其一画上符号。例如,性别分男、女两个答案;婚姻状况分未婚、同居、已婚、分居、离婚、丧偶六个答案。
- ②有序定性回答:列出不同程度的答案,供调查对象选择其一画上符号。例如,您的睡眠好吗? 1表示很好; 2表示好; 3表示一般; 4表示不好; 5表示很不好。
- ③有序定量回答:采用模拟线性评分方法,让调查对象在他们认为适当的线性尺度位置上做出标记。例如,您的睡眠好吗?



注: 0表示睡眠很不好; 10表示睡眠很好; x表示睡眠一般。

(3)综合型问卷:既有封闭型问题,又有开放型问题的问卷称为综合型问卷。一份问卷通常以封闭式为主,根据需要适当增加若干个开放型问题。若在一个封闭型问题的基础上增加一个开放型问题,则称这样的问题形式为半开放型问题。

例如,您的民族是①汉族;②回族;③满族;④壮族;⑤其他:_____(请详细说明)。

4. 问卷制定步骤

- (1)确定研究目的。设计问卷是为了更好地收集调查者所需的信息,所以在设计问卷前首先要确定和明确本研究的研究目的,把握调查的目的和要求,以便调查问卷的内容和调查目的一致。
- (2)提出调查项目形成项目池。问卷设计组成员根据调查目的、内容,查阅有关文献和参考其他调查使用的问卷,提出问卷包含的项目。也可召开有关专家和被调查者参与的小组讨论,提出有关条目。将各人提出的项目汇总并进行整理分析,对含义相同但表达不同者进行统一描述,所有不同的项目构成调查项目池。
- (3)项目筛选。对项目池中的项目采用专家咨询评分、小组讨论等方法进行分析及筛选,以便精简调查项目。也可以通过对预调查的资料进行统计分析来筛选项目。

(4)确定每个项目的提问形式和类型。问卷提问形式和类型可划分为封闭式问题、开放式问题、综合型问题。根据研究目的、项目答案特点、各答案设置类型的优缺点，确定每个条目的提问形式及答案设置类型。

(5)确定每个项目的回答选项。问答选项与项目的提问方式和类型有关。不同的问答题有不同的答案设计方法。常用的有以下几种：

①填空式。例如在调查经济状况时，对每月工资收入需准确了解具体数字，则可采用填空式，如“您的每月工资收入是_____元”。

②是否式。例如回答是否长期吸烟者(每日1支以上，持续吸烟3个月以上)的答案可以是：1表示是，2表示否。

③选择式。如封闭型问题，应列出各种可能的答案，并按一定顺序编号。一般来说，等级项目按等级关系编号，如文化程度；分类项目按照习惯或一定的逻辑顺序编号，如职业可以是：1表示工人，2表示农民，3表示干部等。

④矩阵式或表格式。如多个问题的答案选择相同，可将多个问题列成表格，方便选择。例如在调查健康对日常活动影响程度时，可设计如下格式。

问题	有很多限制	有一点限制	根本没有限制
1.手提日常用品(买菜、购物)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.适度活动(扫地、搬桌子、做操)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.上楼梯(5层楼)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

(6)预调查及初步考评。初步形成的问卷需进行预调查和初步考评。

①将问卷分别送给该研究领域的专家、研究人员以及典型的被调查者，根据他们的经验和认识对问卷进行评论，并指出不妥之处。

②从正式调查的总体中抽取一个小样本，对其进行预调变，然后检查和分析预调查的结果，从中发现问题并进行修改。

(7)修改完善。在上述基础上修改完善，形成最终的问卷。

5. 问卷设计中应注意的问题

(1)问卷说明要简单明了。问卷的说明关系到调查的质量与效果，一般要用委婉或感人的语气，说明调查的目的和意义，尤其是要让被调查者觉得调查对自己有作用和意义，或者能够帮助别人，激发他们积极主动参与调查的积极性。

(2)避免采用含义不确切的词汇、副词和形容词，如“很久”、“经常”、“一些”等，各人理解往往不同，在问卷设计中应避免或减少使用。例如，对于“您是否经常生病?”回答者不知“经常”是指一周、一个月还是一年生一次或多次病；可以改问：“您上月生了几次病?”或者“最近半年内您生了几次病?”

(3)避免断定性问题。例如，“您一天抽多少支烟?”这种问题为断定性问题，被调查者如果根本不抽烟，就会造成无法回答。正确的处理办法是在此问题前加一条“过滤性”问题，如“您抽烟吗?”如果回答“是”，可继续提问，否则就终止提问。

(4)避免引导性提问。引导性提问指所提出的问题暗示出研究者的观点和见解，有使被访者跟随这种倾向回答的可能。例如，“有人认为被动吸烟会导致肺癌，您同意吗?”引导性提问会导致被调查者不加思考就同意问题中暗示的结论。引导性提问是调查的大忌，常常会引出有严重偏倚的结论。

(5)避免被访者难堪和禁忌的敏感问题，包括各地风俗和民族习惯中忌讳的问题、涉及个人利害关系的问题、个人隐私问题等。例如，“您是否有婚外性关系?有过多少次?”。对于这类问题，

被访者往往出于本能的自卫心理,不愿意回答或不予真实问答,而且还会引起反感。如果有些问题非调查不可,可采用敏感问题调查法。

(6)避免笼统、抽象或不确切的问题(年龄、经济收入、家庭人口等),对容易误解的概念应明确限定。调查项目可以产生歧义的理解。例如,年龄有虚岁、实岁;收入是仅指工资,还是包括奖金、补贴、其他收入、实物发放折款收入在内;家庭人口有常住人口和生活在一起的人口。

(7)避免一问多答的问题。一个项目最好只问一个要点,一个项目中如果包含过多询问内容,会使被访者无从回答,给统计处理也带来困难。例如,“您的父母是知识分子吗?”这个问题使那些父母中仅有一个是知识分子的人无法回答“是”或“否”。此外,知识分子本身也需要界定清楚,否则难以回答。

(8)注意提问的顺序。在设计问卷时,要注意提问的顺序,使问卷条理清楚,提高回答效果。可参考以下几点:①被访者容易回答且较为关心的问题应先提问,专业性强的具体细节问题和敏感性应问题应尽量放在后面;②提问的内容应从简单逐步向复杂深化,对相关关联的内容应进行系统的整理,使被访者不断增加兴趣;③封闭性问题应放在前面,开放性应问题应放在后面。

(9)关于定量指标的半定量化:对于有些定量指标,如年龄、经济收入等,如果能调查到具体的数据,最好按定量指标进行调查,便于分析和归类;可将其半定量化。例如,“您的月工资是:①800元以下;②800元以上;③1500元以上;④3000元及以上”。此时,要注意划分的档次不宜太多,各档的数字之间应正好衔接,无重叠、中断现象。

6. 调查表和问卷的评价

设计一份问卷并不难,难的是设计一份好的问卷。问卷的考评一般包括效度、信度、可接受性等。

(1)效度:指问卷的有效性和正确性,也即准确度。意指问卷确实测定了打算测定的特征(而不是其他特征)以及测定的准确程度。一个问卷的效度越高,说明问卷的结果越能反映其所测对象的真正特征。

(2)信度:指问卷测量结果的可靠性、稳定性和一致性,也即精确度。信度一般指同一种测定方法对同一对象重复测量,随机误差引起测量结果的变异程度。这里要注意,量表的评价常用内部信度指标,即问卷各项目间相关性考评。

(3)可接受性:指被调查者对问卷的接受程度。再好的问卷,如果调查者不愿意接受,也难于实行。问卷的可接受性主要取决于以下几个因素:①问卷简洁明了,条目少且容易理解;②问卷内容为被调查者所熟悉,认为有意义(与其生活及健康相关);③问卷容易填写,看完简短的“填表说明”后,被访者可以自己完成填写;④完成问卷所需的时间较少,一般认为5~30min较适宜。临床使用的问卷最好在15min内,一般人群评价的问卷可稍长,但也不宜超过30min,否则被访者会感到厌烦而不愿完成,或随意乱填。具体考察时可通过接受率(问卷回收率)、问卷合格率和填表所需平均时间等来评价。

(4)区分度:指不同个体问答该问题的答案应该有所不同,该项反映个体间的差异。例如在“新鲜蔬菜摄入与食管癌的关系”的调查中,调查的问题是:你是否吃新鲜蔬菜?如果被调查者填写答案都是“是”,这个项目就没有区分度。

其中,信度与效度的考评方法特别适合于各条目均有得分的问卷,如心理测量、态度测量、生存质量测量等的标准化量表。在统计学中有一系列反映和分析信度与效度的指标和方法,详见相关文献。

5.1.5 内科调查设计样例展示及对差错的辨析与释疑

【例5-1】研究者欲现场调查某市区社区居民对糖尿病的认知情况。首先,确定调查对象和观察单位,调查对象是指该市区的所有社区居民,观察单位是该市区每个社区的居民。其次,确定

调查项目、设计调查表,调查项目包含被调查者的基本信息,如姓名、性别、住址、联系方式等以便当调查信息不完整时能联系到被调查者对量表进行补充;根据调查社区居民对糖尿病认知的研究目的设定与糖尿病认知相关的分析项目。

在设定调查项目过程中容易犯的主要错误如下:

(1)调查项目内容包罗万象,不够精简,存在重复。调查条目全固然好,但是如果不够精简的话,不仅造成资源的浪费,还会浪费调查员和被调查员的宝贵时间,更重要的是使完成量表和问卷的时间延长,造成被调查者的厌倦情绪,导致调查偏倚的发生。

(2)调查项目界定不够准确,如“您每天喝酒多吗?”没有对饮酒的量进行具体的界定。应该具体到酒的种类(红酒,白酒还是葡萄酒)及酒大体的度数和饮酒的数量(一般以两为度量衡单位)。

(3)存在一题多问现象,如“您认为吸烟喝酒对糖尿病有害吗?”对该问题详细推敲就会发现,到底是问受试者吸烟对糖尿病的危害,还是饮酒对糖尿病的危害,还是吸烟和饮酒两者对糖尿病的危害。

(4)调查项目采用专业术语,不够通俗易懂,如“您是否知道糖尿病的典型症状是‘三多一少’症状?”“三多一少”过于专业,应该用“多尿、多饮、多食、消瘦”来表达,或是更通俗的说法“小便多、喝水多、吃得多,但是还变瘦”。以上是常见的一些错误,在实际调查过程中尚存在不少其他易犯错误,研究者应在实践中不断琢磨和体会,需要角色换位思考地界定每个调查项目。

设计好调查表之后,最后可以采用随机方法抽取部分该市区社区居民并对他们以现场问卷的方式进行调查。在抽取部分观察单位过程中容易犯的错误是,没有采取严格随机方法进行抽取,而是采用随意的方式选取部分社区居民进行调查。比如在调查过程中,由于样本含量比较大时,进行随机抽样编码等比较麻烦,不少调查者采用随意选择调查对象,或者采用遇到谁调查谁的方法,或者认识或者熟悉哪个小区或者街道办,为了减轻调查难度就按照自己的主观意愿选择这些小区或者街道办居民进行调查的错误方法。这些方法都是在实践过程中经常遇到的一些错误做法,研究者务必在调查过程中杜绝类似问题的发生。

5.2 横断面调查研究的设计与分析

横断面调查研究(cross-section study)是指在特定的时点或时期内,对某一人群的健康状况,以及相关的致病因子、卫生服务状况、环境状况等进行调查,又称现况调查。调查时间不宜过长,一般控制在一个月内较为适当。其目的主要是:

- (1)描述疾病或健康状况的分布情况。
- (2)描述某些因素或特征与疾病的关联,确定危险因素。
- (3)为评价防治措施及其效果提供有价值的信息。
- (4)查出某一地区患有研究疾病的人群,从而达到早发现、早诊断和早治疗的目的。

5.2.1 调查方法

根据调查的范围和调查对象的选择方式可分为普查、抽样调查和典型调查。

(1)普查(overall survey):也称全面调查(complete survey),将组成总体的所有观察单位全部加以调查(如我国人口普查)。理论上,只有普查才能得到总体参数,因为普查没有抽样误差,也不需要假设检验,但因普查工作量大、费时费力,且系统误差大、调查质量难以保证等缺点,除非十分必要,一般不宜采用。

(2)抽样调查(sample survey):是一种非全面调查,它是医学调查研究中最常用的一种方法。抽样调查是随机从总体中抽取一定数量的有代表性的观察单位组成样本,然后根据样本信息推断

总体特征。其优点：省时省力、正确率较高，许多医学问题只能做抽样调查，如药物疗效观察等。缺点：调查设计、实施及统计分析比较复杂。

(3) 典型调查 (typical survey)：也称案例调查，即在对事物进行全面分析的基础上，选择少数典型的人或单位进行深入调查。典型调查没有贯穿随机抽样的原则，不宜进行统计推断，但在一定条件下，结合专业知识，可以对总体特征做出经验推论。

5.2.2 常用的抽样方法

抽样必须遵循“随机化”原则，才能从总体中随机抽取有代表性的观察单位组成样本。根据抽样样本的方法，抽样方法可以分为概率抽样和非概率抽样两类。概率抽样是按照概率论和数理统计的原理从调查研究的总体中根据随机原则来抽选样本，并从数量上对总体的某些特征做出估计推断，对推断出可能出现的误差可以从概率意义上加以控制。

1. 常用概率抽样方法

常用的概率抽样方法有简单随机抽样、系统抽样、整群抽样、分层抽样四种。

1) 简单随机抽样

简单随机抽样 (simple random sampling) 要求每个观察单位有同等概率被选入样本。将调查总体全部观察单位编号，再用抽签法或随机数字表随机抽取部分观察单位组成样本。

一种简单的方法是，从 N 个观察单位的总体中抽取例数为 n 的样本。可先将调查对象的全部观察单位编号并各赋一个随机数 (位数与 N 的位数相同)。再将 N 个随机数排序，序号 $1 \sim n$ 所对应的观察单位为抽取的样本。

【例 5-2】 要想了解某地 3000 名小学生中蛔虫的感染率，拟用抽样方法调查 300 人，可先对该地在校 3000 名小学生每人都编一个号，并做成签，充分混匀后从中随机抽出 300 个签，与这 300 个签相对应的学生组成有代表性的样本。也可利用计算机或计算器产生随机数字。对于上例，将 3000 名小学生编号，从计算机产生多于 300 个随机数字，变换为 3000 以内的数字，如计算机产生 4 位数的随机数为 8619, 7650, 1716, 1818, …, 原随机数在 $1 \sim 3000$ 之内者不变, $3000 \sim 6000$ 之间者减去 3000, $6000 \sim 9000$ 之内者减去 6000, 9000 以上者废弃。这样，前 300 个互不相关的随机数字所对应的小学生就是要调查的对象。

简单随机抽样可以采用掷硬币、掷骰子、抽签或者查随机数字表的方法，也可以在计算机上利用 SAS、SPSS 等软件来实现。

基本步骤：编号→赋予随机数→随机数排序→抽取样本。

例如，从 N 个观察单位的总体中随机抽取例数为 n 的样本，首先将总体中的 N 个观察单位进行编号，其次对每个观察单位赋予一个随机数字，再将赋予的 N 个随机数字按照由小到大进行排序，序号 R 为 $1 \sim n$ 所对应的观察对象为抽取的 n 个样本。

这种抽样的优点是简单直观、均数 (或比率) 及标准误的计算简便；缺点是总体较大时，难以对总体中的个体一一编号，且样本分散，不易组织调查。

【例 5-3】 以表 5-2 为例说明如何实现简单随机抽样。表 5-2 是 20 例病人的基本信息，要求用简单随机抽样的方法从中随机抽取 10 名患者作为调查对象。

表 5-2 20 例病人的基本信息

编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
性别	F	F	M	F	F	F	M	M	M	M	M	M	F	F	F	F	F	M	F	F
年龄	60	64	37	57	41	31	60	64	58	16	58	63	23	37	20	33	39	40	49	42

(1)查随机数字表法：从随机数字表中任意位置(例如第3行第1列)，横向(或纵向)依次获取20个随机数字，遇到相同的随机数字跳过取下一个随机数字，将选取的20个随机数字按照由小到大进行排序，序号为R，则序号R为1~10对应的病人为随机抽取的样本。

(2)用SAS实现单纯随机抽样的程序名为nkksysj5_3.SAS。

```
data nkksysj5_3;
input id sex $ age;
cards;
1      F      60
2      F      64
...    ...    ...
20     F      42
run;
proc surveyselect  method=srs n=10;
run;
```

程序说明：首先，创建SAS数据集名为nkksysj5_3。然后，调用PROC SURVEYSELECT过程进行随机抽样，并指定一些抽样选项；“method=”用于指定随机抽样的方法，“srs”表示单纯随机抽样；“n=”用于指定抽样大小。

简单随机抽样样本均数、样本率标准误的计算公式见表5-3。

表 5-3 简单随机抽样样本均数、样本率标准误的计算公式

总 体 类 型	均数标准误	率的标准误
无限总体	$S_{\bar{x}} = \frac{S}{\sqrt{n}}$	$S_p = \sqrt{\frac{p(1-p)}{n-1}}$
有限总体	$S_{\bar{x}} = \frac{S}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$	$S_p = \sqrt{\frac{p(1-p)}{n-1}} \sqrt{1 - \frac{n}{N}}$

2)系统抽样

系统抽样(systematic sampling)又称等距抽样或机械抽样，是按照某种顺序将总体中的各个体编号，最后随机地抽取某一个体作为第一个体，最后机械地依次每隔若干个体抽取一个个体组成样本。抽取的样本编号为： $i, i+K, i+2K, i+3K, \cdots, i+(n-1)K$ 。其中， i 为随机抽取的第一个编号， K 为抽样间隔， n 为样本量。其方法是按照一定顺序，机械地每隔若干个观察单位抽取一个观察单位组成样本。

【例 5-4】 流行病学调查，要从1000户中抽取10%做样本，可先在门牌号1~10号之间随机抽取一户(假定为第5号住户)，其后每隔10号抽取一户，即抽取5, 15, 25, 35, …, 995，共100户组成样本。

系统抽样的优点是易于理解，简便易行，容易得到一个有代表性的样本。缺点是一般情况下，虽然系统抽样的抽样误差小于简单随机抽样，但是当总体的观察个体按顺序有周期趋势或单调增(或减)趋势时，容易抽取明显偏倚的样本。例如在上述流行病学调查例子中，我们抽取的住户均为单号，其住房可能都是一个朝向，若做采光或者温度、湿度等卫生学检查，将产生明显的偏性。

【例 5-5】 要在1000名新生中抽取200人做样本，其抽样间隔为1000/200=5；若随机抽取的第一号为2，则抽取的个体号依次为2, 7, 12, 17, 22, 27, …。

【例 5-6】 将表5-3中20例病人以编号为序，采用系统抽样方法抽取5名受试者。

(1)查表法：总体N=20例，样本量n=5，因此，抽样间隔K=N/n=20/5=4。从随机数字表

中任意位置(例如第 3 行第 4 列), 横向(或纵向)获取的第一个 01 ~ 05 之间的随机数, 如为 02, 则抽样取的个体号依次为 2, 6, 10, 14, 18。

(2)用 SAS 实现系统随机抽样的程序名为 nkksysj5_6. SAS。

```
proc surveyselect method=sys n=5;
control id;
run;
```

程序说明: 数据步与 nkksysj5_3. SAS 中相同。SURVEYSELECT 过程与 nkksysj5_3. SAS 的不同之处在于选项 method = sys, 规定抽样方法为系统随机抽样法。此外, 还多了一个 control 语句, 该语句用于指定排序的变量。

系统抽样的抽样误差: 总体的性质不同, 抽样间隔不同, 其抽样误差也不同, 所以系统抽样本身无统一的计算标准误的方法。在实际工作中, 一般按照简单随机抽样方法估计误差, 但系统抽样抽取每个观察单位并不是彼此独立的。如上例中, 在 1 ~ 10 之间随机确定 5 后, 15, 25, 35, …便不是独立的。而按简单随机抽样方法估计抽样误差一般是偏大的。

3) 整群抽样

整群抽样(cluster sampling)是先将总体分成 K 个群, 每个群中包含若干个观察单位, 再从 K 个群中随机抽取 k 个群(k < K), 将抽中的 k 个群中的全部个体组成样本, 称为整群抽样。例如调查某县农村儿童贫血状况, 可按乡镇分群体; 再随机抽取几个乡镇; 对其中的全部儿童进行调查。

【例 5-7】 拟研究某地中学生中近视眼的患病率, 从该地 4 所中学中随机地抽取 2 所中学, 对被抽取的 2 所中学的全部中学生进行调查, 即整群抽样。如果从被抽取的 2 所中学内再各随机抽取 2 个班, 对被抽取的班内全部中学生调查, 就是两阶段整群抽样。由于这种抽样方法易于组织, 常在大规模的调查中采用。但各群间一般差异较大, 因而抽样误差也较大。

整群抽样的优点是便于组织, 节省经费, 容易控制调查质量; 缺点是样本例数一定时, 一般抽样误差大于单纯随机抽样。

整群抽样的抽样误差: 设 N 为总体观察单位数, $\sum X$ 为样本中各群全部观察值之和, $\sum m_i$ 为调查各群个体数之和, \bar{X}_i 为样本第 i 群的均数, T_i 为第 i 群内观察值之和, \bar{T}_i 为各 T_i 的均数, $\sum \alpha_i$ 为样本中各群阳性数之和, $\bar{\alpha}$ 为样本各群的平均阳性数, p_i 为样本第 i 群的率。样本均数(\bar{X})、样本率(p)及其标准误的计算公式见表 5-4 和表 5-5。

表 5-4 各群内观察单位数相等时的计算公式

总体类别	样本均数 \bar{X}	均数标准误 $S_{\bar{X}}$	样本率 p	率的标准误 S_p
无限总体	$\frac{\sum X}{\sum m_i} = \frac{\sum \bar{X}_i}{k}$	$\sqrt{\frac{\sum (\bar{X}_i - \bar{X})^2}{k(k-1)}}$	$\frac{\sum \alpha_i}{\sum m_i} = \frac{\sum p_i}{k}$	$\sqrt{\frac{\sum (p_i - p)^2}{k(k-1)}}$
有限总体	$\frac{\sum X}{\sum m_i} = \frac{\sum \bar{X}_i}{k}$	$\sqrt{\frac{\sum (\bar{X}_i - \bar{X})^2}{k(k-1)} \left(1 - \frac{k}{K}\right)}$	$\frac{\sum \alpha_i}{\sum m_i} = \frac{\sum p_i}{k}$	$\sqrt{\frac{\sum (p_i - p)^2}{k(k-1)} \left(1 - \frac{k}{K}\right)}$

表 5-5 各群内观察单位数不相等时的计算公式

总体类别	样本均数 \bar{X}	均数标准误 $S_{\bar{X}}$	样本率 p	率的标准误 S_p
无限总体	$\frac{K}{Nk} \sum m_i \bar{X}_i$	$\frac{K}{N} \sqrt{\frac{\sum (T_i - \bar{T}_i)^2}{k(k-1)}}$	$\frac{K \sum \alpha_i}{Nk}$	$\frac{K}{N} \sqrt{\frac{\sum (\alpha_i - \bar{\alpha})^2}{k(k-1)}}$

续表

总体类别	样本均数 \bar{X}	均数标准误 $S_{\bar{X}}$	样本率 p	率的标准误 S_p
有限总体	$\frac{K}{Nk} \sum m_i \bar{X}_i$	$\frac{K}{N} \sqrt{\frac{\sum (T_i - \bar{T}_i)^2}{k(k-1)} \left(1 - \frac{k}{K}\right)}$	$\frac{K \sum \alpha_i}{Nk}$	$\frac{K}{N} \sqrt{\frac{\sum (\alpha_i - \alpha)^2}{k(k-1)} \left(1 - \frac{k}{K}\right)}$

【解】 阳性率 $P = \frac{\sum \alpha_i}{km} = \frac{12 + 17 + \cdots + 18}{8 \times 50} = 0.34$,

$S_p = \left\{ \left(1 - \frac{8}{80}\right) \times \frac{1}{8 \times (8-1)} [(0.24 - 0.34)^2 + (0.34 - 0.34)^2 + \cdots + (0.36 - 0.34)^2] \right\}^{\frac{1}{2}}$
= 0.0248

总体率的 95% 可信区间为 $0.34 \pm 1.96 \times 0.0248$, 即 (0.291 , 0.389) 。

4) 分层抽样

分层抽样 (stratified sampling) 先按对观察指标影响较大的某种特征, 将总体分为若干个类别 (统计上称之为“层”, strata), 再从每一层内随机抽取一定数量的观察单位, 合起来组成样本。

【例 5-9】 如调查某县农村妇女下生殖道感染情况, 可按乡镇经济发展水平分层 (如好、中、差三层), 再在各层中对各乡镇做随机抽样。如调查某地肺癌死亡率, 由于不同年龄段的肺癌死亡率不同, 故按年龄组进行分层抽样, 这里, 对观察指标——肺癌死亡率影响较大的特征即年龄。如分为 0 ~ 14 岁组、15 ~ 34 岁组、35 ~ 54 岁组、55 岁及以上组, 再从每个年龄组中随机抽取一定数量的观察单位合起来组成样本。在分层时, 应当使样本中各层的比例接近总体的比例, 如分别按总体中年龄构成比例确定每个年龄组的调查人数, 这样就增强了样本的代表性。

分层抽样的方式一般有按比例抽样和最优抽样两种方案。

(1) 按比例分配 (proportional allocation): 按各层观察单位数 N_i 占总体观察单位数 N 比例抽取样本, 使各层样本含量 n_i 与样本总含量 n 之比等于各层观察单位数 N_i 与总体观察单位数 N 之比。采用按比例分层随机抽样时, 所得均数或比例是自动加权的。样本量分配可按下式计算:

$$\frac{n_i}{n} = \frac{N_i}{N} \text{ 或 } n_i = N_i \frac{n}{N}$$

【例 5-10】 在 12 万人口的居民区中调查某病患病率。居民区分 4 层, 样本总含量确定为 1000 人, 根据按比例分配的计算方法, 问: 各层应抽取多少人?

根据公式 $n_i = N_i \frac{n}{N}$ 计算可得, 各层应分别抽取 292、208、417 和 83 人, 合计 1000 人, 如表 5-6 所示。

表 5-6 各层应抽取人数的分配表

层	人口数 (N_i)	抽样比例 (N_i/N)	样本含量 (n_i)
1	35000	0.292	292
2	25000	0.208	208
3	50000	0.417	417
4	10000	0.083	83
合计	120000 (N)		1000 (n)

(2) 最优分配 (optimum allocation): 同时按总体各层观察单位数 N_i 的多少和标准差 σ_i 的大小分配, 按下面两式分配各层的样本量, 使抽样误差最小。

均数的抽样公式: $n_i = n \frac{N_i \sigma_i}{\sum N_i \sigma_i}$

率的抽样公式：
$$n_i = n \frac{N_i \sqrt{\pi_i(1 - \pi_i)}}{\sum N_i \sqrt{\pi_i(1 - \pi_i)}}$$

式中, σ_i 为总体第 i 层的标准差(参数), π_i 为总体第 i 层的率(参数), σ_i 或 π_i 一般根据以往经验、文献资料或预调查来估计。

分层随机抽样的优点是合理地设计分层可以使样本具有较好的代表性, 减少抽样误差, 可以对不同层进行独立分析; 但分层不合理也可使样本的代表性欠佳。

【例 5-11】 对表 5-6 中的病人以性别为分层因素, 等比例抽取 10 名受试者。

(1) 查表法: 将这 20 名患者按性别进行分层, 再用单纯随机抽样法或系统抽样的方法从各层中抽取患者。共有 12 名女性、8 名男性, 因此应从女性中抽取 6 名受试者, 从男性中抽取 4 名受试者。从男女受试者中再随机抽取相应个数的受试者, 方法同前。

(2) 用 SAS 实现分层随机抽样的程序名为 nkkysj5_11.SAS。

```
proc sort data = nkkysj5_11;
by sex;
run;
proc surveyselect method = srs n = (6 4);
strata sex;
run;
```

程序说明: 数据步与 nkkysj5_3.SAS 中相同。过程步 SORT 过程用于将数据集 nkkysj5_11 排序, by 语句指明按性别进行排序, 此处排序方式为先 F 后 M。SURVEYSELECT 过程与 nkkysj5_3.SAS 的不同之处在于选项 n = (6 4), 规定每层中的抽样大小(第 F 层抽 6 例、第 M 层抽 4 例)。此外, 还多了一个 STRATA 语句, 该语句用于指定分层变量。

分层抽样的抽样误差: 样本均数(\bar{X})、样本率(p)及标准误的计算见表 5-7。

表 5-7 分层抽样样本均数、样本率及其标准误的计算公式

总体类别	样本均数 \bar{X}	均数标准误 $S_{\bar{X}}$	样本率 p	率的标准误 S_p
无限总体	$\sum W_i \bar{X}_i$	$\sqrt{\sum W_i^2 S_{\bar{X}_i}^2}$	$\sum W_i p_i$	$\sqrt{\sum W_i^2 S_{p_i}^2}$
有限总体	$\sum W_i \bar{X}_i$	$\sqrt{\sum \left(1 - \frac{n_i}{N_i}\right) W_i^2 S_{\bar{X}_i}^2}$	$\sum W_i p_i$	$\sqrt{\sum \left(1 - \frac{n_i}{N_i}\right) W_i^2 S_{p_i}^2}$

2. 非概率抽样

以上几种抽样方法都考虑了等概率的原则, 即抽样框中的成员或单位被抽中的概率相同。如果对总体无法确知, 只能采用非概率抽样方法, 它主要依据研究者的意愿、判断或方便程度等条件来抽取调查对象, 因不考虑随机抽样中的等概率原则, 往往产生较大的抽样误差, 难以保证样本的代表性。常用的非概率抽样方法有立意抽样、偶遇抽样、滚雪球抽样等。

(1) 立意抽样: 又称为目的抽样或判断抽样, 是研究者根据研究目标和对情况的主观判断来选择和确定调查对象的方法。假如欲调查吸毒者吸毒的原因, 由于吸毒者较少, 吸毒行为又及其隐蔽, 无法判断其总体有多大, 无法采用随机抽样的方法, 只能在吸毒人群较多的地区进行调查, 或选择对吸毒者有档案管理的社区进行。

(2) 偶遇抽样: 又称为便利抽样或方便抽样, 指调查者根据实际情况使用对自己最为便利的方式来选取样本, 可以是抽取偶然遇到的人或选择那些容易找到的人作为调查对象。例如, 某医生要

调查病人对目前医疗收费的看法，直接选择到该医生科室看病的患者进行调查就是便利抽样。

(3) 滚雪球抽样：当无法了解总体情况时，可以从找到的少数个体入手，对他们进行调查，并请他们介绍其他符合条件的人，扩大调查面，如此重复下去，直到达到所需的样本含量。例如，对吸毒者、女性性服务者的调查可以采用滚雪球抽样的方法。

5.2.3 样本含量的估计

在抽样调查中，要考虑样本大小问题，因为样本例数过少，所得指标不够稳定，检验效能会降低。而样本量过多，不但会造成不必要的人力、物力、财力的浪费，而且也给质量控制带来困难。估计样本例数即在保证一定精度和检验效能前提下，确定最少的观察单位数目。抽样的方法不同，样本含量计算方法不同，下面仅列举简单随机抽样样本含量估计方法。（其他抽样方法调查设计样本含量估计方法请参阅相关统计专业书籍。）

- 估计样本含量的前提条件如下。
- (1) 定出检验水准 α ：首先确定本次实验中允许犯第 I 类错误的概率 α ，通常取 $\alpha = 0.05$ ；同时还应明确是单侧还是双侧检验。
 - (2) 确定所容许误差 δ ：预计样本统计量与总体参数最大差别控制在什么范围，常取可信区间宽度的一半。
 - (3) 总体标准差 σ ：若未知，可根据文献估计，或者通过预试验获得样本标准差 S 作为总体标准差 σ 的估计值进行样本含量的估计。
 - (4) 若对有限总体进行抽样，还需要了解总体观察单位数 N 。
- 当研究者确定 α ，并对 δ 和 σ 做出估计后，就可以根据设计类型和要求估计样本含量。

1. 估计总体均数样本含量

无限总体抽样：
$$n = \left(\frac{u_{\alpha} \sigma}{\delta} \right)^2, \delta = \bar{x} - \mu$$

有限总体抽样：
$$n_c = \frac{n}{1 + \frac{n}{N}}$$

式中， n 为所需的样本含量， α 有单、双侧之分， u_{α} 为相应的正态分位数。

【例 5-12】 某研究者拟用单纯抽样方法了解某地区成年男子血红蛋白的平均水平，希望误差不超过 2g/L，根据文献，血红蛋白的标准差约为 25g/L，如 $\alpha = 0.05$ （双侧）。问：需调查多少病人？

【解】 $u_{0.05} = 1.96$ 代入公式 $n = \left(\frac{1.96 \times 25}{2} \right)^2 = 600$ 。所以，需调查 600 人。

2. 估计总体率样本含量

(1) 当 π 接近 0.5 时，

无限总体抽样：
$$n = \frac{u_{\alpha}^2 \pi (1 - \pi)}{\delta^2}, \delta = p - \pi$$

有限总体抽样：
$$n_c = \frac{n}{1 + \frac{(n - 1)}{N}}$$

【例 5-13】 据以往调查结果，高血压的患病率为 40%，某研究者欲了解某地高血压的患病率，希望误差不超过 2%。问：需调查多少人？

【解】 取 $\alpha = 0.05$ （双侧）， $u_{0.05} = 1.96$ ， $\delta = 0.02$ ， $\pi = 0.4$ ，代入公式得

$$n = \frac{1.96^2 \times 0.4 \times (1 - 0.4)}{0.02^2} = 2305$$
 ; 所以, 需要调查 2305 人。

(2) 当所调查的率偏向两侧时(如 $P < 0.3$ 或 $P > 0.7$), 正态性较差, 往往需要对样本率 P 做以弧度为单位的变换。样本量估计公式为:

无限总体抽样:
$$n = \left(\frac{57.3 u_{\alpha}}{\arcsin \left[\frac{\delta}{\sqrt{\pi(1-\pi)}} \right]} \right)^2$$

有限总体抽样:
$$n_c = \frac{n}{1 + \frac{(n-1)}{N}}$$

5.2.4 资料的整理和分析

资料的整理: 主要包括资料的审核(包括项目审核、数据审核和逻辑审核)和对原始资料用科学的方法按照调查目的进行分类汇总, 使原始资料系统化, 以便准确地反映事实, 使人一目了然, 是研究工作不可缺少的环节。资料整理应遵循完整性、标准性、真实性、准确性及合理分类的原则。

资料统计分析: 包括统计描述和统计推断, 通过计算相应的指标和进行必要的统计学处理, 结合专业知识, 进而阐明事物的内在规律, 然后得出结论。不同的资料类型采用不同的统计分析方法, 定量资料可以通过算术均数、几何均数、中位数、标准差、四分位数间距等指标进行统计描述; 定性资料可以用率、构成比、相对比、绝对数等指标进行统计描述。此外, 还可以用统计表和统计图等对资料进行描述。统计推断多用 t 检验、方差分析、 u 检验、 χ^2 检验等假设检验方法。

5.2.5 内科横断面调查研究样例展示及对差错的辨析与释疑

【例 5-14】 某杂志拟通过研究了解 35 岁以上湖北省高血压病的流行现状, 为开展高血压人群防治和采取有效的干预措施提供科学依据。调查方法是, 采用多阶段分层随机抽样方法, 在全省随机抽取 10 个市县, 根据地理分布确定 3 个农村点、3 个城镇点、4 个城市点。城市点抽取 1 条街道、城镇点抽取 1 个城镇、农村点抽取 1 个乡镇。对抽取街道或乡镇 35 岁以上常住居民开展调查, 每个点调查 1200~1500 人, 如符合条件调查对象不足 1200 人, 在临近街道/乡镇补足。该研究实际上采用的是多阶段分层随机整群抽样方法, 杂志中提到采用的多阶段分层随机抽样方法有些不妥。后面提到的符合条件调查对象不足 1200 人, 在临近街道/乡镇补足, 既然是采用的分层随机整群抽样, 人数不足以在临近街道/乡镇补足, 有点画蛇添足。

本研究欲调查 35 岁以上居民高血压患病率在男、女之间有无统计学差异, 调查结果见表 5-8。

表 5-8 35 岁以上居民高血压患病率在男、女性别间的比较

性 别	患高血压例数	未患高血压例数	合 计 例 数
男性	1762	5043	6805
女性	1723	5337	7060
合计	3485	10380	13865

分析与计算: 因研究两个总体率之间有无差别, 故可采用横断面研究设计 2×2 表资料的 χ^2 检验。下面以表 5-8 为例进行分析。SAS 实现程序见 nkksysj5_14. SAS)。

SAS 程序如下:

```
Data nkkysj5_14;
Do A =1 to 2;
Do B = 1 to 2;
Input f @@;
Output;
End;
End;
Cards;
1762      5043
1723      5337
;
Proc freq;
Weight f;
Tables A* B/chisq;
Run;
```

输出结果如下：

Statistics for Table of A by B

Statistic	DF	Value	Prob
Chi-Square	1	4.0751	0.0435
Likelihood Ratio Chi-Square	1	4.0746	0.0435
Continuity Adj. Chi-Square	1	3.9964	0.0456
Mantel-Haenszel Chi-Square	1	4.0748	0.0435
Phi Coefficient		0.0171	
Contingency Coefficient		0.0171	
Cramer's V		0.0171	
Fisher's Exact Test			
Cell (1,1) Frequency (F)	1762		
Left - sided Pr <= F	0.9792		
Right - sided Pr >= F			
0.0228			

统计和专业结论；由于 $\chi^2=4.0751$ ， $P=0.0435<0.05$ ，表明男女性别间高血压患病率有统计学差异，35 岁以上年龄段中男性高血压患病率高于女性患病率。

说明：SAS 分析结果全部给出各种统计分析的结果。若手工计算，当总频数 $n>40$ 且无理论频数小于 5 时，可选用计算简便的一般 χ^2 检验；当总频数 $n>40$ 但至少有一个理论频数 $1<T<5$ 时，可选用校正的 χ^2 检验；当总频数 $n<40$ 或至少有一个理论频数 $T<1$ 时，必须选用 Fisher 的精确概率检验。

5.3 追踪性调查研究的设计与分析

5.3.1 基本概念

追踪性调查研究称为队列研究 (cohort study)，又称为随访研究 (follow-up study) 或纵向研究，是检验疾病病因假说的一种重要的研究方法，属于“由因寻果”的前瞻性研究，故也称为前瞻

性调查研究。在队列研究中，“队列”泛指共同暴露于某一因素或者具有某种共同特征的一组人群，即指研究对象接触过某种物质(如放射线、接种某种疫苗等)、具备某种特征(年龄、性别、种族、职业、遗传特性等)、具有某种行为或习惯(抽烟、经常晨练)等。暴露在不同的研究中有不同的含义，暴露可以是有害的，也可以是有益的，也可能是与疾病无关的。

队列研究的基本原理是，从总体中选择和确定所需的研究对象，根据目前或过去某个时期是否暴露或其暴露的不同水平，将研究对象分成不同的组，例如分成暴露组和非暴露组或分成低剂量、中剂量或高剂量组。各组除暴露因素外，其他方面的条件应基本接近。对各群组中的所有观察对象均随访一定时期，观察并记录在这个时期内所研究的结局(如发病，死亡等)，比较两个群组在观察期内该结局的发生率，判断研究因素与疾病间存在的统计学联系及联系程度，从而进一步论断暴露因素与疾病的联系，暴露组与非暴露组的发病率计算如表 5-9 所示。

队列研究的主要观察指标是相对危险度(relative risk, RR)，为暴露组的发病率(或死亡率)与非暴露组的发病率(或死亡率)之比： $RR = \frac{P_1}{P_0}$ ，RR 表示暴露组人群相对于非暴露组人群发病危险性的大小。RR 越大，表明暴露导致人群发病的危险性越大， $RR > 1$ 表示暴露因素是疾病的危险因素， $RR < 1$ 表示暴露因素是疾病的保护因素， $RR = 1$ 表示暴露与疾病无联系。

表 5-9 追踪性调查研究(队列研究)资料整理表

组 别	发 病 人 数	观 察 人 数	发 病 率
暴露组	a	N_1	$P_1 = \frac{a}{N_1}$
非暴露组	b	N_0	$P_0 = \frac{b}{N_0}$
合计	x	N	$P = \frac{x}{N}$

5.3.2 追踪性调查研究的类型

1. 前瞻性队列研究(prospective cohort study)

前瞻性队列研究是队列研究的基本形式。研究对象的分组是根据研究对象现时的暴露情况而定的，此时研究的结局还没有出现，需要前瞻性观察一段时间才能得到结局。

前瞻性队列研究的特点是，依据暴露因素来分组，观察开始时，事件未出现，需要追踪观察一定时期，方能得到结果(发病或死亡)，故其性质是前瞻性的，即从现在追踪到将来。

前瞻性队列研究的最大优点是，研究者可以直接获取第一手资料，而且资料的偏性比较小。但是如果需要观察大量人群，则花费太大；如果疾病的潜伏期很长，则需要观察的时间很长，这些都会影响队列研究的可行性。

2. 回顾性队列研究(retrospective cohort study)

回顾性队列研究又称为历史性队列研究，根据研究开始时研究者已经掌握的有关研究，对研究对象过去某个时点的暴露状况的历史资料对研究对象进行分组。虽然研究是现在开始的，但是研究开始时，结局已经出现，研究对象是在过去某个时点进入队列的，不需要前瞻性观察。

回顾性队列研究是依据过去某时点的暴露因素分组，然后从已有记录中追溯从那时起到现在为止，队列成员中的死亡或发病情况。研究开始时，某事件已经发生。

回顾性队列研究与前瞻性队列研究相比，节省人力、物力、时间，出结果快，因而适用于长诱

导期和长潜伏期的疾病，也经常用于对特殊暴露人群的研究。但对暴露的测量较为困难，常常缺乏影响暴露与疾病关系的混杂因素的资料，以致影响暴露组和非暴露组的可比性。

3. 双向性队列研究 (ambispective cohort study)

双向性队列研究即在历史性队列研究基础上，继续前瞻性观察一段时间，是前瞻性队列研究与历史性队列研究结合起来的一种设计模式。

当有关暴露情况的资料在过去已经收集到或从现有的记录中可以查到，但疾病的结局尚不知道或不完全了解时，研究人员须随访研究对象的疾病发生情况，以获得疾病结局的确切资料。双向性队列研究兼有上述两类的优点，且在一定程度上弥补了两者的不足之处。

5.3.3 研究对象的选择

1. 暴露人群的选择

暴露人群可简单地分为特殊暴露人群和一般暴露人群两类。

(1)特殊暴露人群：是具有某种特殊暴露经历的人群，例如暴露于污染饮用水的居民，暴露于病毒感染的婴幼儿。如果暴露因素与疾病有关，则高度暴露的人群中疾病的发病率和死亡率就可能高于其他人群，这将有利于探索暴露与疾病之间的联系。所以，在研究暴露与疾病的关联时，常常首先选择特殊暴露人群。由于某些职业常存在特殊暴露因素，所以某些职业人群也是常用的特殊暴露人群。例如选择染料厂工人研究联苯胺导致膀胱癌的作用，选择石棉作业工人研究石棉与肺癌的关系等。

(2)一般暴露人群：一般暴露人群由具有不同暴露因素的个体组成。通过对一般暴露人群进行队列研究，可以探讨某些暴露因素与疾病或健康的关系，适用于观察多种暴露和多种疾病之间的关系。例如选择某癌肿高发县全部居民作为研究对象，研究的暴露因素可以同时包括饮用污染水、黄曲霉毒素摄入量、肝炎等。这种队列的不利之处是，通常必须面访每个研究对象来获得暴露信息。选择研究队列通常从两点考虑：①所研究的因素与疾病是人群中常见的；②计划观察一般人群的发病情况，特别是观察环境因素与疾病的关系，无特殊暴露人群或不需要特殊暴露人群。

2. 暴露人群的选择

非暴露人群由没有受到暴露因素影响的个体组成，作为队列研究中暴露人群的对照，选择对照组的目的是进行比较。研究结果真实性依赖于是否正确地选择了非暴露人群对照。因此，要注意与暴露人群的可比性，对照人群除了未暴露于所研究的因素外，其他各种因素或人群特征（如年龄、性别、职业、文化程度等）应尽可能地与暴露人群相似。选择对照时，一般有三种方式：

(1)内对照：同一研究人群中采用没有暴露或暴露水平最低的人员作为对照即为内对照。内对照是比较理想的对照，因为除暴露因素外，与暴露人群的可比性较好；同时选用内对照较为方便、可行。Doll 与 Hill 进行的“吸烟与肺癌”的研究以及“Framingham 心脏病研究”均是采用内对照。

(2)外对照：没有暴露或具有较低暴露水平的其他人群属于外对照。职业人群或特殊暴露人群常需在该人群之外特设对照组。如把放射科医生作为研究放射线致病的暴露人群，则可选用不接触射线或接触射线较少的五官科医生作为外对照。

(3)一般人群对照：一般人群的发病率或死亡率比较稳定且较容易得到。用一般人群作为对照可以节省研究经费和时间。它的缺点是资料比较粗糙，往往不十分精确或缺乏要比较的项目。特别要注意的是，应当使用暴露人群所在地区的人口，而且尽可能保证两组在性别、年龄、民族、经济水平、调查时间等方面具有可比性。

5.3.4 样本含量的估计

追踪研究一般需要抽取一定数量的样本进行研究。样本大小的估计应当考虑下面两个问题：(1)暴露组与对照组的比例；(2)队列研究的失访率。估计样本时要考虑到失访率，通常按 10% 来估计失访率，即按计算出来的样本量再加 10% 作为实际样本量。

追踪研究中的样本量大小取决于四个参数：(1)非暴露人群的发病率 p_0 ；(2)暴露人群的发病率 p_1 ；(3)检验水准 α ，通常 α 取 0.05 或 0.01；(4)所要达到的把握度，即 $1 - \beta$ ，为拒绝无效假设的能力或避免假阴性的能力。

1. 暴露组和对照组样本量相等时

用以下公式计算出各组所需的样本含量：

$$n = \frac{(z_{\alpha} \sqrt{2\bar{p}\bar{q}} + z_{\beta} \sqrt{p_0q_0 + p_1q_1})^2}{(p_1 - p_0)^2}$$

其中， p_1 与 p_0 分别表示暴露组和对照组的预期发病率， \bar{p} 为两个发病率的平均值。

$\bar{p} = (p_0 + p_1)/2$ ， $\bar{q} = 1 - \bar{p}$ ， $q_0 = 1 - p_0$ ， $q_1 = 1 - p_1$ ， Z_{α} 、 Z_{β} 分别为标准正态分布下的面积。

【例 5-15】某医师采用队列研究的方法评价某药物预防脑卒中再发的效果，得知不用药者脑卒中的再发概率为 23%，估计 RR 值为 0.5，设 $\alpha = 0.05$ ， $\beta = 0.10$ 。问：需要多大样本量？

【解】已知 $\alpha = 0.05$ ， $\beta = 0.10$ ， $Z_{0.05} = 1.960$ ， $Z_{0.10} = 1.282$ ， $p_0 = 0.23$ ， $q_0 = 1 - p_0 = 0.77$ ， $p_1 = RR \times p_0 = 0.5 \times 0.23 = 0.115$ ， $q_1 = 1 - p_1 = 0.885$ ， $\bar{p} = (0.23 + 0.115)/2 = 0.173$ ， $\bar{q} = 1 - \bar{p} = 0.827$ 。

代入公式得 $n = 225.12$ ，即用药组和非用药组各需要 226 例。

【例 5-16】某队列研究观察放射线暴露者与白血病的关系，已知普通人群白血病发病率 1/万，放射线暴露者发病率为 10/万。若 α 取为 0.05， β 取为 0.1。问：各需观察多少人？

【解】本例已知， $p_0 = 0.0001$ ， $p_e = 0.001$ ， $u_{\alpha} = u_{0.05} = 1.645$ ， $u_{\beta} = u_{0.1} = 1.282$ （单侧），

$\bar{p} = (p_e + p_0)/2 = (0.0001 + 0.001)/2 = 0.00055$ ，代入公式得到：

$$n_e = n_0 = \frac{2 \times 0.00055(1 - 0.00055)(1.645 + 1.282)^2}{(0.001 - 0.0001)^2} = \frac{0.00941887}{0.00000081} = 11628$$

故两组各需观察 11628 人。

2. 暴露组 n 与非暴露组 n_0 不相等时

若 $n:n_0 = 1:c$ ，

$$n = \frac{(z_{\alpha} \sqrt{2\bar{p}\bar{q}} + z_{\beta} \sqrt{p_0q_0 + p_1q_1})^2}{(p_1 - p_0)^2}, n_0 = c \times n, N = n + n_0$$

实际工作中应考虑失访所造成的样本量减少，在估计样本含量时应给予适当补偿。

5.3.5 资料的整理和分析

分析资料前，首先应该对资料进行审核，对有明显错误的资料进行重新调查、修正或剔除；对不完整资料要设法补齐。在此基础上一般先做描述性分析，对研究对象的组成、人口学特征、随访的经过、时间长短、结局的发生和失访的情况等做出描述。然后进行统计推断分析，分析暴露组和非暴露组两组率的差异，通过计算相对危险度与假设检验，推断暴露的效应和大小。追踪性调查研究（队列研究）统计分析资料的基本模式如表 5-10 所示。

表 5-10 追踪性调查研究(队列研究)统计分析资料的基本模式

组 别	病 例	非 病 例 组	合 计	发 病 率
暴露组	a	b	$a + b = n_1$	$P_1 = \frac{a}{n_1}$
非暴露组	c	d	$c + d = n_0$	$P_0 = \frac{b}{n_0}$
合计	$a + c = m_1$	$b + d = m_0$	$a + b + c + d = N$	$P = \frac{x}{N}$

1. 人时的计算

因为队列研究时间跨度较长,观察对象处于动态之中,队列内对象被观察的时间可能很不一致,因此以人为单位计算率就不合理。较合理的办法是加入时间因素,以“人时”来计算观察对象的暴露经历。

(1)以个体为单位计算暴露人年。

该方法结果准确,但计算费事;如果样本不大,可用此法。现在已经有专门用于计算人年的计算机软件,如 PYRS、OCMAP 等。

(2)近似法计算暴露人年。

可应用近似法计算暴露人年,即用平均人数乘以观察年数得到总人年数,平均人数一般取相邻两年的年初人口的平均数或年中人口数。该法计算简单,但精确性较差。

(3)寿命表法计算人年。

当观察人数较多时,利用寿命表法计算人年较为方便。下式计算中,当年进入队列的人按 0.5 人年计算,失访或出现终点结局的也按 0.5 人年计算。

$$L_x = I_x + 1/2(N_x - D_x - W_x), I_{x+1} = I_x + N_x - D_x - W_x$$

式中, L_x 为 x 时间内暴露人年数, I_x 为 x 时间开始的人数, N_x 为 x 时间内进入队列的人数, D_x 为 x 时间内出现终点结局的人数, W_x 为 x 时间内失访的人数。

2. 率的计算

(1) 累计发病率(cumulative incidence, CI)。

当观察人口比较稳定时,无论观察时间长短,以开始观察时的人口数为分母,整个观察期内发病人数为分子,计算该观察期的累计发病率(cumulative incidence, CI);用同样的方法可计算累计死亡率。计算公式为 $q = d/n$; 式中 q 表示累计发病(死亡)率, d 表示观察期内新发病例(或死亡)数, n 表示观察总人数。如有 1000 名工人同时暴露在粉尘作业环境下,在 20 年的追踪观察中,共查出尘肺新发病人 200 人,则 20 年的累计尘肺发病率为 $q = 200/1000 = 0.2$ (或 20%)。

(2) 发病密度(incidence density, ID)。

发病密度是指当观察的人口不稳定,观察对象进入研究的时间先后不一,以及各种原因造成失访时,每个观察对象随访的时间不同,若用总人数为单位计算率是不合理的。此时可用人时(person-time)为单位计算发病率或死亡率,即以观察人数乘以观察时间。用人时为单位计算的率带有瞬时频率性质,称为发病密度。时间单位可用年、月、日、时等,最常用的是年,即以人年(person-year)为单位计算率。每人暴露一年就是 1 人年,1 人暴露 10 年或 10 人暴露 1 年都计算为 10 人年,人年发病率的计算公式为 $f = d/T$, 式中 f 为人年发病率, T 是观察总人年数, d 是观察期内的发病人数。如上例观察的 1000 人中,假设有 150 人观察到 10 年时、另 200 人观察到 15 年时出现了失访现象,则研究中观察的总人年数为 $150 \times 10 + 200 \times 15 + (1000 - 150 - 200) \times 20 = 17500$, 累积尘肺发病率 $q = 200/17500 = 0.0114$ 人年或 114/万人年,即平均每 1 万人年发生 114 个尘肺病人。

(3) 标化比(standardized mortality ratio, SMR)。

当研究对象数目较少, 结局事件的发生率比较低时, 无论观察的时间长或短, 都不宜直接计算率, 而是以全人口发病(死亡)率作为标准, 算出该观察人群的理论发病(死亡)人数, 即预期发病(死亡)人数, 再求观察人群实际发病(死亡)人数与此预期发病(死亡)人数之比, 得到标化发病(死亡)比。最常用的指标为标化死亡比(standardized mortality ratio, SMR), 这一指标在职业病流行病学研究中常用。

SMR 虽然是在特殊情况下用来替代率的指标, 但实际上不是率, 而是以全人口的发病(死亡)率作为对照组计算出来的比。

【例 5-17】 某厂 30~40 岁年龄组工人有 500 名, 在某年内有 2 人死于肺癌, 已知该年全人口 30~40 岁组肺癌的死亡率 2‰, 求其 SMR。

【解】
$$\text{SMR} = \frac{\text{研究人群中的察死亡}(O)}{\text{以准人口死亡率算出的期死亡}(E)}$$

已知 $O=2$, $E=500 \times 2‰=1$, $\text{SMR}=2/1=2$

即某厂 30~40 岁年龄组工人死于肺癌的危险达到相应一般人群的 2 倍。

(4) 率的假设检验。

由于队列研究多为抽样研究, 当发现两组率有差别时, 首先要考虑抽样误差导致的可能, 需要进行统计学假设检验; 常用 u 检验。

当研究样本量较大, p 和 $1-p$ 都不太小, 如 np 和 $n(1-p)$ 均大于 5 时, 样本率的频数分布近似正态分布, 此时可应用正态分布的原理来检验率的差异是否有统计学意义, 即用 u 检验法来检验暴露组与对照组之间率的差异。

$$u = \frac{p_1 - p_0}{\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_0}\right)}}$$

式中, p_1 为暴露组的率, p_0 为对照组的率, n_1 为暴露组观察人数, n_0 为对照组的观察人数, p_c 为合并样本率, $p_c = \frac{X_1 + X_0}{n_1 + n_0}$, 其中 X_1 和 X_0 分别为暴露组和对照组结局事件的发生数。求出 u 值后, 查 u 界值表得 P 值, 按所取的检验水准即可做出判断。

3. 暴露因素与疾病的关联强度指标

(1) 相对危险度(relative risk, RR)。

相对危险度也叫危险度比(risk ratio, RR), 为暴露组发病(或死亡)率 p_e 与非暴露组发病(或死亡)率 p_0 之比, 是队列研究中用于描述某因素与发病(或死亡)之间关联的主要统计学指标。 $\text{RR} = p_e/p_0$

RR 表明暴露组发病或死亡的危险是对照组的多少倍。RR 值越大, 表明暴露的效应越大, 暴露与结局关联的强度越大, $\text{RR}=1$ 表示研究因素与发病(或死亡)无关; $\text{RR}>1$ 说明研究因素可能为危险因素; $\text{RR}<1$ 说明研究因素可能为保护因素。

RR 的检验如下。

① 成组设计资料通常用 Mantel-Haenszel χ^2 检验: $\chi^2 = \frac{(ad-bc)^2 n}{n_1 n_0 m_1 m_0}$ 。

② 配对设计资料对通常用配对设计 χ^2 检验: $\chi^2 = \frac{(b-c)^2}{b+c}$ 。

RR 的 95% 可信区间常用 Miettinen 法: $\text{RR}^{(1 \pm 1.96/\sqrt{\chi^2})}$ 。

(2) 归因危险度(attributable risk, AR)。

归因危险度又叫特异危险度、危险度差(risk difference, RD)和超额危险度(excess risk)。归因危险度是暴露组发病率与对照组发病率之差,反映暴露因素引起的发病率的改变量。用发病率或死亡率计算归因危险度的公式为: $AR = p_e - p_0 = p_0(RR - 1)$, 其中 p_e, p_0 分别为暴露和非暴露组的发病率(或死亡率)。

(3) 归因危险度百分比(attributable risk percent, AR%)。

归因危险度用百分比表示时称为归因危险度百分比, 又称病因分值(etiological fraction, EF), 指暴露人群中归因于暴露的发病和死亡占全部发病或死亡的百分比。其计算公式为:

$$AR\% = \frac{p_e - p_0}{p_e} \times 100\% = \frac{RR - 1}{RR} \times 100\%$$

4. 剂量反应关系分析

如果某种暴露存在剂量反应关系,即暴露的剂量越大,其效应越大,则该种暴露作为病因的可能性就越大。分析方法是先列出不同暴露水平下的发病率,然后以最低暴露水平组为对照,计算各暴露水平的相对危险度和危险度差。必要时,应对危险度(或率)的变化做趋势性检验。剂量反应关系反映暴露与疾病之间的共变关系,由表 5-11 可以看出血清胆固醇水平与冠心病之间的共变关系,即随着观察人群血清胆固醇水平的升高,其患冠心病的相对危险度也在升高。剂量反应关系是对暴露因素与疾病进行因果推论的重要依据之一。

表 5-11 40~59 岁男子按初始血清胆固醇水平分组的冠心病 6 年发病情况

血清胆固醇 (mg/dl)	人 数	病 例 数	危 险 度	年均发病率	RR
<210	454	16	0.0352	0.0059	1.00
210 ~	455	29	0.0637	0.0106	1.81
≥245	424	51	0.1203	0.0200	3.39
合计	1333	96	0.0720	0.0120	—

5. 分层分析与多变量分析

单因素分析方法有时难以适应病因研究对统计分析提出的要求。分层分析是把人群根据某特征分为不同层次,如按性别可以分为男、女,按年龄可分为不同年龄组,然后分别分析各层中暴露与疾病的关联。因为分层的因素是可能的混杂因素,通过分层可以控制这些因素的干扰。logistic 回归和 cox 回归等技术可以进行多因素分析,对于控制因素的混杂作用有明显的优越性,可以探讨疾病的危险因素、混杂因素以及研究因素之间的交互作用,也可以估计不同暴露水平下个体患病的可能性。详见第 14 章、第 15 章中的有关内容。

5.3.6 追踪性调查研究中的偏倚及控制

1. 常见的偏倚

(1) 选择偏倚: 在研究过程中,由于某些原来选择的研究对象拒绝参加或另选他人代替,或在进行历史性队列研究中某些人的记录丢失或不完全,或不同亚组的应答率不同,都会导致不同暴露组成员中主要特征分布不一致,从而出现选择偏倚。

(2) 随访偏倚: 在随访期间,暴露或非暴露组中总会有些研究对象由于种种原因脱离了观察,无法继续随访他们。主要原因有研究对象迁移、外出、不愿再合作而退出或死于非终点疾病,这些情况称为失访(loss of follow-up)。队列研究由于观察人数多、观察时间较长,失访偏倚是不可避免的,但失访率一般不应超过 10%。

(3)信息偏倚:信息偏倚主要为错分偏倚(misclassification bias),包括暴露错分、疾病错分以及暴露与疾病的联合错分。主要原因是使用的仪器不准确,检验技术不熟练,诊断标准定义不明确或掌握不当,询问技巧欠佳造成结果不真实等。若这种错分偏倚以同样的程度发生于观察的各组,则结果可能不会对各组之间的相对关系产生太大影响,否则,将影响到相对危险度的估计。

(4)混杂偏倚:当我们研究某个因素与某种疾病的关联时,由于某个既与疾病有制约关系,又与所研究的暴露因素有联系的外来因素的影响,掩盖或夸大了研究的暴露因素与疾病的联系,该外来因素所带来的影响称为混杂偏倚。

2. 偏倚的控制

仔细、严格地按规定的标准选择研究对象以及尽力弄清应答者和未应答者之间的差别,就可减少选择偏倚。如有差别,研究人员虽然无法消除偏倚,但也有助于他们了解因所使用的选择方法带来的偏倚方向和程度。运用精确的测量和规定的分类标准就能减少信息偏倚,当然应当知道确定这些分类标准的方法的灵敏度和特异度。尽量随访全部对象并提高他们的依从性就可以减少失访偏倚,同时还应同样对待每个研究对象,即当他们出现所研究疾病的临床特征和体征时均按规定的标准予以诊断,还要比较失访者和未失访者基线调查的特征,这样就可以提供信息以估计由于失访所造成的偏倚的性质和程度。在设计时利用限制或匹配的方法,资料分析阶段采用分层分析或多因素分析模型处理,可适当控制混杂因素。

5.3.7 追踪研究的优缺点

1. 优点

(1)追踪研究能追踪疾病发展过程与死亡的发生,能对暴露因素及其他影响因素进行监测,因而所得资料可靠、准确。

(2)可以直接获得不同队列的发病和死亡资料,直接计算反映暴露因子与发病或死亡的关联和关联强度的指标RR、AR。所以理论上说,队列研究能初步检验疾病与病因之间的因果关系。

(3)可以观察并获得多种疾病的结局资料,可同时研究一种暴露因素与多种疾病的关系。

2. 缺点

(1)前瞻性队列研究方法适用于研究常见病,所需的研究对象数量较大。

(2)常常需要较长时间地观察(随访)研究对象,尤其对于诱导期较长的慢性病。因此,难以保持良好的依从性,容易产生失访偏倚,而且研究对象的各项特征包括暴露与否都可能发生改变。

(3)计算暴露人年的任务较繁重。

(4)开展一项队列研究比较费时、费钱,且需要大量人力,准备与组织工作都较为复杂和繁重。

由于追踪研究的设计要求严密,分析资料时也要求具有较高的技术,因此,实际工作中一般先开展病例对照研究,在获得意义比较明确的暴露因素后,再进行队列研究。

5.3.8 内科追踪性调查研究样例展示及对差错的辨析与释疑

【例5-18】《中国新生儿科杂志》刊登了一篇文献拟探讨早产儿支气管肺发育不良(BPD)的发生率及危险因素。研究方法采用前瞻性队列研究的方法,对来某医院产科的2009年1月—2010年6月出生的所有活产早产儿进行研究,统计BPD发生率,并分析相关危险因素。

在进行追踪研究选择研究对象时要有确定的入选标准和排除标准。本研究入选标准为:2009年1月—2010年6月在本院出生的早产儿(胎龄<37周)。排除标准为:(1)出生时已为死胎;

(2)在产房，因家属放弃抢救治疗而死亡；(3)入住新生儿重症监护室(NICU)后因病情危重或社会经济原因放弃治疗而死亡。

前瞻性追踪研究要对结局有一个确定的界定，如本研究中早产儿 BPD 诊断标准采用美国国家儿童保健和人类发展研究院(NICHHD)通过的 BPD 诊断标准：(1)早产儿或低体重儿，出生后 28 天(胎龄 32 周者)或校正胎龄 36 周(胎龄 <32 周者)仍需持续氧疗或机械通气；(2)患儿有慢性或进行性呼吸功能不全，表现为低氧血症或高碳酸血症；(3)胸部 X 线片及肺部 CT 表现为两肺不同程度的肺纹理增多模糊或毛玻璃影囊泡形成线状及网格状影；(4)排除先天性心脏病、气胸、胸腔积液、膈疝等疾患。

通过调查：(1)孕妇住院时填写妊娠及孕期患病情况；(2)记录新生儿分娩方式、胎龄、性别、出生体重、多胎、宫内窘迫、Apgar 评分、有无使用 PS 等；(3)每周观察两次入住新生儿病室和 NICU 患儿的症状及体征，记录治疗措施和实验室及其他辅助检查结果，吸氧早产儿根据病情每周或 10 天拍胸部 X 线片 1 次；(4)以患儿确诊 BPD 或出院或死亡为调查结束时间，即队列研究的终点，进行前瞻性追踪调查研究。

统计分析方法：以 $P < 0.05$ 为差异有统计学意义作为标准，采用单因素分析应用 Pearson ChiSquare 分析或 Fisher 确切概率法，筛选出潜在的危险因素；在单因素筛选出危险因素基础上，再进行多因素 logistic 回归分析筛选出早产儿 BPD 发病的危险因素。Pearson Chi-Square 分析的 SAS 程序与前 cross_section. sas 程序相似，logistic 回归分析 SAS 程序应用 proc logistic 命令，在此不再赘述。

5.4 回顾性调查研究的设计与分析

5.4.1 基本概念

回顾性调查研究(respective study)，又称病例—对照研究(case-control study)是一种“由果溯因”的回顾性研究。其基本原理是以确诊患某病的一组病人作为病例组，以不患有该疾病但具有可比性的另一组个体为对照组，通过回顾性调查两组在过去的某段时间对某种危险因素(研究因素)的暴露情况，比较病例组和对照组各因素的暴露率及暴露水平差异，判断研究因素与疾病间是否存在统计学联系及联系程度，进一步推断该暴露因素与疾病的联系。回顾性调查研究(病例—对照研究)的资料应该整理成如表 5-12 所示的形式。

可见，病例—对照研究是：(1)在疾病发生后进行的，已经有一批可供选择的病例；(2)研究对象按发病与否分成病例组与对照组；(3)调查的被研究因素的暴露情况是由研究对象从现在对过去的回顾；(4)从因果关系的角度看，结果已经发生，是由果推因；(5)只能了解两组的暴露率或暴露水平。

表 5-12 回顾性调查研究(病例对照研究)资料整理表

组 别	暴 露 数	非 暴 露 数	合 计 数
病例组	a	b	$a + b$
对照组	c	d	$c + d$
合计	$a + c$	$b + d$	n

在病例对照研究中，不能获得发病率资料，无法估计相对危险度。但是，可以用比数比(Odd Ratio, OR)表示疾病与暴露的联系。 a/b 表示病例中暴露与非暴露的比例， c/d 表示对照中暴露与非暴露的比例，则比数比 OR 定义为：

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

显然, OR 越大, 表明暴露导致人群发病的可能性越大。OR > 1 表示暴露因素是疾病的危险因素, OR < 1 表示暴露因素是疾病的保护因素, OR = 1 表示暴露与疾病无联系。病例—对照研究主要应用于罕见疾病、“潜伏期”较长的疾病的病因学研究, 与队列研究相比, 节省成本且容易实施。可得到与结局有关的多个病因因素的资料, 是探索可疑影响因素的快捷途径。

5.4.2 回顾性调查研究的类型

1. 病例和对照不匹配

病例和对照不匹配的回顾性研究也称“成组”设计病例对照研究, 指在设计所规定的病例和对照人群中, 分别抽取一定量的研究对象进行研究。一般对照的数目等于或多于病例人数, 此外没有其他任何特别限制与规定。

2. 病例和对照匹配

匹配(matching)或称配比, 即要求对照在某些因素或特征与病例保持一致, 目的是对两组进行比较时排除匹配因素的干扰。如以年龄作为匹配因素, 在分析比较两组资料时, 可避免由于两组年龄构成的差别对疾病和因素的影响, 从而更正确地说明所研究因素与疾病的关系。匹配分为频数匹配和个体匹配。

(1) 频数匹配。

频数匹配(frequency matching)是指使匹配的因素在对照组与在病例组所占的比例一致。如病例组中男女各半, 65 岁以上者占 1/3, 则对照组中也是如此。

(2) 个体匹配。

个体匹配(individual matching)以病例和对照的个体为单位进行匹配称为个体匹配。1:1 匹配称为配对(pair matching), 1:2, 1:3, ..., 1:m(m > 1)称为匹配。

采用匹配的目的在于提高研究效率(study efficiency)。匹配后, 用匹配因素来做分层分析, 可使各层的病例和对照分布均衡, 从而使其效率更高。匹配的特征或变量是已知的混杂因子, 或有充分的理由怀疑为混杂因子, 否则不应匹配。

3. 其他衍生的研究类型

近年来, 衍生出一些非上述传统的病例对照研究类型, 主要有巢式病例对照研究、病例队列研究、病例交叉研究等研究类型。巢式病例对照研究(nested case-control study)又称嵌入式回顾性研究, 是将队列研究和病例对照研究相结合的一种研究方法。首先根据科研设计收集研究队列中每个成员的暴露信息以及有关的混杂资料, 确认随访期内发生的每个病例; 然后以队列中的病例作为病例组, 对照组来自同一队列, 进行病例对照研究。该方法的优点是: ①病例与对照的暴露资料均在发病和死亡之前获得, 暴露与疾病的时间先后顺序清楚, 而且不存在回忆偏倚; ②病例组和对照组可比性好; ③可提高统计效率和检验效率, 因为队列成员的暴露率较高, 而且队列成员都有共同的开始暴露时间, 而回顾性研究只取整个暴露期的一个横断面。其他衍生类型在内科学上用得较少, 请参阅专业统计书籍或流行病学书籍。

5.4.3 病例与对照的选择

病例与对照的选择是设计中一个重要的问题, 其基本原则是调查的病例足以代表总体中该病的病例, 调查的对照足以代表产生病例的总体。

1. 病例的选择

(1)病例的定义。病例的定义受到诊断标准、疾病严重程度、诊断标准是主观还是客观等因素的影响。对诊断标准不明确界定,就会产生错误分类,应尽量选用国际通用且统一的诊断标准进行疾病定义。

(2)病例的来源。主要来源有两种:一种以医院为基础;另一种以地区或人群为基础。

以医疗保健机构为基础选择病例:如从医院某一特定人群中获得全部确诊的病例,此法较易进行、省经费,所以是最常用的选择病例的方法。此时应选某一段时间内的全部病例,诊断必须明确,诊断方法和诊断标准统一,对病期和病型也要事先规定。由于病人对医疗机构具有选择性,某一医院的病例不能代表全人群该病特点。

以地区或人群为基础选择病例:如以某地区某段时间内的全部病例作为调查对象,此法较为妥当,选择对照较简单,不易产生选择偏倚。但实际工作中不能经常包括全部病例。其优点是所得结果可以推广到研究人群中。

无论是哪种来源都要有严格的统一的标准,并尽可能包括全部病例,最好用新发病例而不用患病者,因前者对暴露情况的回忆比后者好,能尽量避免回忆偏倚。

(3)病例的类型。主要有三个类型:新发病例、现患病例、死亡病例。最好用新发病例而不用现患病例,因前者对暴露情况的回忆比后者好,能尽量避免回忆偏倚。

2. 对照的选择

设置对照组的作用在于平衡除研究因素外的其他可能影响发病与否的因素。所以对照组的选择应该与病例组具有很好的均衡性,通常要求两组除所要研究因素外,其他各方面的特征(如年龄、性别、职业、经济状况、种族、饮食习惯等)都应尽可能一致。对照可以是健康人,也可以是患其他病的病例,但不能是同一系统疾病的患者,如结核病与慢性支气管炎都与吸烟有关,两者不能互为对照。

从医院中抽取对照:必须从与可疑因素无关的各种疾病的所有病人中随机取得,要明确不包括哪些病人,不要把该病的可疑病人选入。由于医院病例易有偏性,容易出现选择偏倚,因此,一般选用配对法选取对照。

从地区或人群抽取对照:如果病例是精确确定的人群中所有病例的一个代表性样本,对照则可直接从人群中抽取。如果有户口登记或能够对人群进行编码,可用完全随机抽样法选取对照。

5.4.4 样本含量的估计

在两事物间关系的调查中,病例—对照研究是常用的方法。常用的病例—对照研究中,可以是不配对的,也可以是配对的;可以是病例数和对照数相等的,也可以是不相等的。下面介绍在不同情况下所需的样本含量估计方法。

1. 不匹配但病例数与对照数相等时

各组需要的样本含量:

$$n = \frac{(z_{\alpha} \sqrt{2\bar{p}\bar{q}} + z_{\beta} \sqrt{2p_0q_0 + p_1q_1})^2}{(p_1 - p_0)^2}$$

式中, p_0 表示对照组中暴露者的比例, $q_0 = 1 - p_0$; p_1 表示病例组中暴露者的比例,

$$q_1 = 1 - p_1; \bar{p} = \frac{1}{2}(p_1 + p_0), \bar{q} = 1 - \bar{p}$$

[注]: 如果无 p_1 的估计值, 但有备择假设中的优势比的估计值 OR, 则可用下式计算 p_1 :

$$p_1 = \frac{p_0 \cdot \text{OR}}{1 + p_0(\text{OR} - 1)}$$

式中, OR 应取优势比 OR 的可能取值中的最小值。

【例 5-19】 在饮酒与心肌梗死病例对照的研究中, 已知对照组饮酒者的比例为 20%, 比值比 OR=2.5, 定 α 为 0.05, β 为 0.10。问: 需调查多少病例与对照?

【解】 已知 $p_0=0.2$, $\text{OR}=2.5$, 有

$$p_1 = \frac{p_0 \text{OR}}{1 + p_0(\text{OR} - 1)} = \frac{0.2 \times 2.5}{1 + 0.2(2.5 - 1)} = 0.3846$$

$$\bar{p} = (p_1 + p_2)/2 = (0.2 + 0.3846) = 0.2932$$

$u_\alpha = u_{0.05} = 1.645$ (单侧), $u_\beta = u_{0.1} = 1.282$ (单侧), 代入公式

$$n_1 = n_2 = \frac{2 \times 0.2932 \times (1 - 0.2932) (1.645 + 1.282)^2}{(0.3846 - 0.2)^2} = 104$$

故这项研究工作需调查病例和对照各 104 名。

2. 不匹配且病例数与对照数不等时

若设病例组例数与对照组例数之比为 1:c, 则病例组的样本含量可按式计算:

$$n = \frac{(z_\alpha \sqrt{2\bar{p}\bar{q}} + z_\beta \sqrt{2p_0q_0 + p_1q_1})^2}{(p_1 - p_0)^2}$$

式中, $\bar{p} = \frac{p_1 + cp_0}{1 + c}$, $\bar{q} = 1 - \bar{p}$ 。

3. 1:1 配对设计的病例—对照组研究

调查结果有 4 种情况, 即“++”, “--”, “+-”, “-+”。其中, “++”和“--”称为一致结果, “+-”和“-+”称为不一致结果。配对调查的优点是可以对某些干扰因素的作用作均衡处理, 并可能减少样本含量。其样本含量 (或总对子数 N) 的计算公式为:

$$N \approx \frac{n}{p_0(1 - p_1) + p_1(1 - p_0)}$$

式中, $n = \left(\frac{u_\alpha/2 + u_\beta \sqrt{p(1-p)}}{p - 0.5} \right)^2$, $p \approx \text{OR}/(1 + \text{OR})$, p_1 为估计的病例中暴露者的比例; p_0 为估计的对照组中暴露者的比例; n 为不一致的对子数。

【例 5-20】 在饮酒与心肌梗死病例对照的研究中, 已知对照组饮酒者的比例为 20%, 比值比 OR=2.5, 定 α 为 0.05, β 为 0.10, 若进行 1:1 配对设计, 需调查多少病例与对照?

【解】 $p = \frac{2.5}{1 + 2.5} = 0.7143$, 代入公式

$$n = \left(\frac{1.645/2 + 1.282 \times \sqrt{0.7143(1 - 0.7143)}}{0.7143 - 0.5} \right)^2 = 42.8$$

$$N \approx \frac{42.8}{0.2 \times (1 - 0.3846) + 0.3846(1 - 0.2)} = 99$$

即需 99 对病例与对照, 所需样本含量略小于不配对时的样本含量。

5.4.5 资料的分析

1. 统计学描述

(1) 描述研究对象的一般特征: 描述研究对象的人数以及各种特征的构成, 例如性别、年龄、

职业、出生地、居住地、疾病类型的分布等。使用频数匹配时应描述匹配因素的频数比例。

(2)均衡性检验：比较病例组和对照组某些基本特征是否相似或齐同，目的是检验病例组与对照组的可比性。对确有统计学意义的因素，在分析时应考虑到它对其他因素的可能影响。

2. 统计学推断

病例对照研究中，由于病例和对照最多只是所有病例和对照的有代表性的样本，并不知道暴露组和未暴露组观察各是多少，故无法直接计算真实的发病率，也无法直接计算相对危险度，但可估计相对危险度。

(1)成组资料的分析。

最简单的病例对照研究所得结果可归纳为如下表 5-13 所示的格式。

表 5-13 回顾性调查研究(病例对照研究)统计分析基本格式

	暴 露 数	无 暴 露 数	合 计 数
病例组	a	b	m_1
对照组	c	d	m_0
合计	n_1	n_0	n

①比数比(odds ration, OR)：由于在病例对照中无法计算相对危险度，在发病率很低的情况下，通常用比数比 OR 来估计它。 $OR = \frac{ad}{bc}$ 。

②OR 的检验：由于通过随机样本计算出的 OR 具有变异性，应排除抽样误差的影响后才能对疾病与暴露之间的关系做出统计学结论，对比数比检验的零假设为 $H_0: RR = 1$ ，对立假设 $H_1: RR \neq 1$ ，检验统计量用 Mantel-Haenszel 提出的公式。

$$H_0: RR = 1, H_1: RR \neq 1,$$
$$\chi^2 = \frac{(ad - bc)^2 n}{n_1 n_0 m_1 m_0}, \nu = 1$$

③OR 的 95% 可信区间(Miettinen 法)： $OR^{(1 \pm 1.96/\sqrt{\chi^2})}$ 。

(2)配对资料的分析。

配对资料是由病例与对照结合成对子，分析结果时不应把对子拆开分析，1:1 配对资料常可整理成如表 5-14 所示的形式。

表 5-14 1:1 配对设计病例对照研究统计分析基本格式

病 例	例 数		对 子 数
	对照：有暴露	无 暴 露	
有暴露	a	b	$a + b$
无暴露	c	d	$c + d$
对子数	$a + c$	$b + d$	n

①OR 在配对条件下的比数比计算公式为： $OR = \frac{b}{c}$ 。

②OR 的检验： $H_0: RR = 1, H_1: RR \neq 1$ ，所用统计量是 $\chi^2 = \frac{(b - c)^2}{(b + c)^2}$ 。

对子数较少时，用校正公式 $\chi^2 = \frac{(|b - c| - 1)^2}{(b + c)^2}, \nu = 1$ 。

③OR 的 95% 可信区间(Miettinen)法： $OR^{(1 \pm 1.96/\sqrt{\chi^2})}$ 。

在配对病例对照研究中,除1:1设计外,还有1: m 设计,即1个病例与 m 个对照配比, $m>1$;危险因素除按两水平分类之外(有或无、高或低等),还有多水平分类。

(3) 不匹配分层资料分析。

所谓分层分析,就是把暴露和未暴露人群或者病例与对照放在较为均衡的条件下进行比较。由于每层之中混杂因素的变化很小,使得各层中该因素与暴露无联系,因此每层提供的暴露和疾病联系的结果都未被混杂。例如,某项研究中性别可能是混杂因素,我们可以将资料按男性和女性分层,然后分别计算男性和女性中暴露和疾病的联系,从而排除了性别因素的混杂影响,这就是分层分析。各层资料OR值及其95%可信区间的计算与成组资料的分析方法相同,总OR值的计算用Mantel-Haenszle提出的公式。

(4) 多因素分析。

在回顾性研究中往往需要同时研究多个因素的作用;若简单地应用一般分层分析方法且对若干因素加以控制,将是复杂的过程,且受样本含量的限制。目前出现了不少多因素分析软件,使得人们可以比较容易地分析多个因素与疾病的联系、联系程度以及各因素之间的相互关系,如条件和非条件logistic模型、loglinear回归模型等。

5.4.6 回顾性研究中的偏倚及控制

1. 选择偏倚

这种偏倚常发生在研究设计阶段,是指由于选入的研究对象与未选入的研究对象在某些特征上存在差异而引起的误差。

(1)入院率偏倚,也称 Berkson 偏倚。当利用医院病人作为病例和对照时,由于对照是医院的某一部分病人,而不是全体目标人群的一个随机样本;而且病人对医院及医院对病人双方都有选择性,因此病例只是该医院的或某些医院的特定病例,难免产生偏倚。尤其当各种疾病的入院率不同时,将导致病例组和对照组某些特征上存在系统差异。

(2)现患病例—新发病例偏倚,又称奈曼偏倚(neyman bias)。如果调查对象选择现患病例,即存活病例,则可能得到更多的信息,但其中很多信息可能只与存活有关,却未必与该病的发病有关,从而高估了某些暴露因素的病因作用。

(3)检出征候偏倚,也称暴露偏倚(unmasking bias)。病人常因某些与致病无关的症状就医,从而提高了早期病例的检出率,致使过高地估计了暴露程度,产生系统误差。

(4)时间效应偏倚(time effect bias),对于肿瘤、冠心病等慢性病,从开始暴露于危险因素到出现病变往往经历一个较长的过程。因此,在回顾性研究时,那些暴露后即将发生病变的人,或已经发生早期病变而不能检出的人,以及在调查中已有病变但因缺乏早期检测手段而被错误地认为非病例的人,都可能被选入对照组,从而导致对研究因素与疾病关系的错误估计。

2. 信息偏倚

信息偏倚又称观察偏倚或测量偏倚,是指在收集整理信息过程中,由于测量暴露与结局的方法有缺陷造成的系统误差。

(1)回顾性研究主要是调查研究对象既往的暴露情况,由于被调查者记忆失真或不完整造成的系统误差。回忆偏倚的产生与调查时间和事件发生的时间间隔、被调查者的构成以及调查技术有关。病例组与对照组的回忆误差不一样,病例组记忆较为准确,且易提供一些自认为与疾病有关的暴露。

(2)调查偏倚,可能来自于调查对象及其调查环境与条件不同,或者调查质量不高及仪器设

备的问题等均可产生调查偏倚。例如，病例在医院调查，而对照在家里调查；调查者对病例和对照的态度不同，有意无意地诱导调查对象以符合研究的病因假设。

3. 混杂偏倚

回顾性研究中同样会发生混杂偏倚，混杂因素和混杂偏倚的概念及其产生原因同队列研究，参见上节介绍。

4. 偏倚的控制

在设计阶段应尽量采用随机方法选择研究对象、在多个医院选择对象等方法以减少入院率偏倚。调查时明确规定纳入标准，或有可能将暴露程度、暴露时间和暴露结局联系起来做结论可减少奈曼偏倚的程度。在调查中尽量采用敏感的疾病早期检查技术，或进行观察期较长的纵向调查，可以尽可能控制时间效应偏倚。选择不易为人们所忘记的重要指标做调查，并重视问卷的提问方式和调查技术，将有助于减少回忆偏倚。调查中尽量采用客观指征；选择合适的人员参加调查，认真做好调查技术培训，采取复查等方法做好质量控制；检查条件尽量一致，尽量在同一时间内由同一调查员调查病例和对照；使用的检查仪器精良，使用前校准；严格掌握试剂的要求等均可减少调查偏倚。采取配比，分层分析、多元分析等可有效控制混杂偏倚。

5.4.7 回顾性研究的优缺点

1. 优点

- (1) 回顾性研究所需样本量小，病例容易找到，因此工作量小，所需人力、物力较少，易于进行，出结果快。
- (2) 可以同时以一种疾病的多种病因进行研究，适合对病因复杂、发病率低、潜伏期长的疾病进行研究。
- (3) 还可以对疗效做出初步的评价。

2. 缺点

- (1) 回顾性研究需要被调查者回忆过去对某因素的暴露史，难以避免出现回忆偏倚，合理地选择对照较为困难。
- (2) 不能计算暴露与非暴露人群的发病率、死亡率和 RR、AR，只能推算比数比。
- (3) 回顾性研究可以阐明某暴露及某病之间是否存在统计学上的联系，但不易下因果关系的结论。

5.4.8 内科学回顾性调查研究样例展示及对差错的辨析与释疑

【例 5-21】《中国糖尿病杂志》刊载一篇文献，采用以医院为基础的 1:1 配对病例—对照研究方法，探讨 II 型糖尿病 (T2DM) 合并冠心病的危险因素，旨在为制订预防控制 T2DM 合并冠心病的策略和措施提供科学依据。

研究对象采用收集的 100 对病例和对照，男 44 对，女 56 对，病例和对照的选择均是以医院为基础进行的，选择的均为某医院住院的 T2DM 患者。为了排除年龄因素造成的偏倚，应保证两组年龄分布差异无统计学意义。

诊断标准：诊断标准需要统一，DM 的诊断按照 1999 年 WHO 诊断标准，排除与 DM 相关的其他内分泌及肝胆疾病等；冠心病诊断依据临床表现、心电图和心脏超声检查。

病例的选择：为合并有冠心病的 T2DM 患者，且冠心病发生于 T2DM 诊断之后。

对照的选择：选择与病例同性别、同年龄组的配比条件，且不合并冠心病的 T2DM 患者中，按

1:1 配对选取对照。

资料收集：基于流行病学和内科学的专业知识，编制信度、效度均符合要求的调查表，采用直接询问的方法进行调查，内容包括人口统计学资料、DM 现病史、高血压史、其他疾病史、疾病家族史、个人情况、社会心理因素等可能是 T2DM 合并冠心病的危险因素。

资料分析：对变量进行单因素条件 logistic 回归分析，计算各可能影响因素的 OR 值及其 95% CI，结合专业知识对可能的混杂因素进行调整。选择单因素分析及调整后仍有统计学关联的变量进行多因素条件 logistic 回归分析，估计各因素对 T2DM 合并冠心病的联合作用。

【例 5-22】研究者拟采用 1:1 匹配的病例对照设计研究接触苯职业史与白血病的关系，并用 logistic 回归模型分析苯职业接触史对白血病的危险程度。

本研究者拟通过回顾当前的白血病患者病例组和非白血病患者对照组以往的苯职业接触史，探讨苯职业接触史是否是白血病的危险因素，若是，危险因素其危害程度如何，以此可以判定该研究属于回顾性的病例对照研究类型。

首先根据研究目的和要求，根据以往相关文献白血病的患病率、设定检验水准 α 的大小和 β 及允许误差的大小，和 1:1 配对设计类型病例对照研究计算样本含量的计算公式，计算需要收集的病例和对照数量，假定需要 303 对病例和对照。

因白血病在普通人群中患病率不是很高，若从普通人群中收集一定数量的病例相对比较困难，所以该研究适宜采用以医院为基础进行病例和对照的选择。

选择几所综合性大医院收集白血病患者 303 例，然后针对每一个白血病患者病例在性别和年龄大致相仿的前提下，为其选择同一所医院就诊的患有其他疾病但未患白血病的病人作为对照。然后详细询问每一个病例和对照是否有苯职业接触史。按照如表 5-15 所示的表格形式对数据进行记录。假定收集的数据资料如表 5-15 所示。

表 5-15 接触苯职业史与白血病的关系 1:1 匹配研究资料

病例接触苯职业史	例 数		对 子 数
	* :有	无	
有	2	30	32
无	12	259	271
对子数	14	289	303

注：* 代表“对照组接触苯职业史”。资料来源：方积乾. 医学统计学与电脑实验(第二版). 2001。

可以运行如下 SAS 程序，程序见 nkkysj5_22。

```
行号      SAS 语句
01      Data nkkysj5_22;
02      Input x1  x2  f  @@;
03      y=1;
04      x=x1 - x2 ;
05      output;
06      Cards;
07      1 1 2
          1 0 30
          0 1 12
          0 0 259
08      ;
09      Proc logistic ;
10      Model y=x/noint ;
11      Weight F;
12      Run;
```

说明：程序第 2 行 x1 及 x2 分别表示病例与对照的暴露水平；用 1, 0 分别表示暴露与非暴露, f 为相应格子中的频数, 第 3 行产生的一个常数因变量 Y, 第 4 行的 x 用以比较(1, 0)和(0,1)的 OR 估计值。运行 SAS 程序时，必须删掉前面的序号。输出结果见表 5-16。

表 5-16 SAS 统计分析结果

变 量	参数估计值	标 准 误	Wald - χ^2 值	P 值	标准化估计值
X	0.9163	0.3416	7.1965	0.0073	1.865984

表中暴露的参数估计值 =0.9163，其标准误为 0.3416，OR = Exp(0.9163) =2.50，95% 的可信区间为 Exp(0.9163 ±0.3416)，即(1.280,4.883)。

(吕军城)

第 6 章 常用内科科研数据统计分析方法

6.1 如何合理应用统计描述指标

统计分析包括统计描述和统计推断。统计描述是指利用统计图、统计表、统计指标等来描述样本资料的特征，统计描述是统计分析的起点和基础。

6.1.1 统计资料的类型

1. 定量资料

定量资料(quantitative data): 又称为数值变量(numerical variable), 其变量值是定量的, 表现为数值大小, 可经测量取得数值, 多有度量衡单位, 如身高(cm)、体重(kg)、血压(mmHg /kPa)、脉搏(次/min)和白细胞计数等。这种由数值变量的测量值构成的资料为定量资料。大多数的数值变量为连续型变量, 如身高、体重、血压等; 而有的数值变量的测定值只能是正整数, 如脉搏、白细胞计数等。

2. 定性资料

定性资料: 又称为分类资料(categorical data), 其变量值是定性的, 表现为互不相容的类别或属性。分类变量可分为无序变量和有序变量两类。

(1) 无序分类变量(unordered categorical variable) 是指所分类别或属性之间无程度和顺序的关系。它又可分为: ①二项分类, 如性别(男、女)、药物反应(阴性和阳性)等; ②多项分类, 如血型(O、A、B、AB)、职业(工、农、商、学、兵)等。对于无序分类变量的分析, 应先按类别分组, 清点各组的观察单位数, 编制分类变量的频数表, 所得资料为无序分类资料, 也称定性资料。

(2) 有序分类变量(ordinal categorical variable) 各类别之间有程度上的差别。如尿样品中测定尿糖的化验结果按“-、±、+、++、+++”分类; 疗效按“治愈、显效、好转、无效”分类。对于有序分类变量, 应先按等级顺序分组, 清点各组的观察单位个数, 编制有序变量(各等级)的频数表, 所得资料称为有序或等级资料。

但是, 变量类型不是一成不变的, 根据研究目的的需要, 精度高的变量可以转化为精度低的变量。例如血红蛋白量(g/L)原属数值变量, 若按血红蛋白正常与偏低分为两类时, 可按二项分类资料分析; 若按重度贫血、中度贫血、轻度贫血、正常、血红蛋白增高分为五个等级时, 可按等级资料分析。有时也可将有序资料数量化, 如可将病人的恶性反应以0、1、2、3表示, 但仍属于有序资料, 不过, 在要求不高时, 可近似按数值变量资料(定量资料)代入回归方程进行分析(应慎用)。

6.1.2 如何合理应用统计描述指标描述定量资料

描述定量资料分布特征的指标有两类: 一类是描述分布集中趋势的指标, 另一类是描述分布的离散程度的指标。

1. 集中趋势指标描述定量资料

描述集中趋势最常用的统计指标是平均数(average): 包括算术均数、几何均数、中位数等。

1) 算术均数

算术均数简称均数(mean)，是算术均数(arithmetic mean)的简称。常用均数反映一组同质观察值的平均水平，总体算数均数用希腊字母 μ 表示，样本均数用 \bar{x} 表示，样本例数用 n 表示。

(1) 适用资料类型：对称分布、正态分布或近似正态分布的数值变量资料。

(2) 计算：有直接法和加权法两种计算方法。

①直接法：计算公式 $\bar{x} = \frac{\sum x}{n}$ ；式中，希腊字母 Σ (读为sigma)表示 x_1, x_2, \dots, x_n 之和； n 为样本含量，即观察值的个数。

②加权法(weighting method)：计算公式 $\bar{x} = \frac{\sum fx_0}{n}$ ；式中 x_0 为该组段的组中值， f 为所在组段的频数。加权法常用于频数表资料或样本中相同观察值较多时。

2) 几何均数(geometric mean, G)

几何均数： n 个数数据乘积的 n 次方根，称为几何均数。

(1) 适用资料：主要适用于两种类型资料。对数正态分布，即数据经过对数变换后呈正态分布的资料；等比数列资料，即观察值之间呈倍数或近似倍数变化的资料，如医学实践中的抗体滴度、平均效价等。

(2) 计算：有直接法和加权法两种计算方法。

①直接法： $G = \sqrt[n]{x_1x_2\cdots x_n} = \lg^{-1}\left(\frac{\sum \lg x}{n}\right)$ ，式中 \lg 代表取常用对数， \lg^{-1} 代表取常用对数的反数，也可用自然对数 \ln 代替。

【例 6-1】7 名慢性迁延性肝炎患者的 HBsAg 滴度资料为 1:16, 1:32, 1:32, 1:64, 1:64, 1:128, 1:512。试计算其几何均数。

【解】 $G = \sqrt[7]{16 \times 32 \times 32 \times 64 \times 64 \times 128 \times 512} = 64$ 也可以用公式 $G = \log^{-1}\left(\frac{\sum \log X}{n}\right)$ 计算。

②加权法： $G = \log^{-1}\left(\frac{f_1\lg X_1 + f_2\lg X_2 + \cdots + f_k\lg X_k}{f_1 + f_2 + \cdots + f_k}\right) = \log^{-1}\left(\frac{\sum f\lg X_k}{\sum f}\right)$ 。式中， X_k 为第 k 组数据的组中值， f 为第 k 组的频数。加权法常用于分析频数表资料。

3) 中位数(median, M)

中位数是一组由小到大按顺序排列、位次处在中间位置的数值。

(1) 适用资料：中位数可用于描述非正态分布资料(对数正态分布除外)；频数分布的一端或两端无确切数据的资料，即开口资料；总体分布不清楚的资料。

(2) 计算。

①直接法：首先将观察值从小到大排列，其次用以下公式计算。

数值个数为偶数时， M 为中间两个数的平均值，即 $M = \frac{1}{2}(X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)})$ 。

数值个数为奇数时， M 为位次居中的数值，即 $M = X_{(\frac{n+1}{2})}$ 。

式中， $X_{(\frac{n+1}{2})}$ 、 $X_{(\frac{n}{2})}$ 、 $X_{(\frac{n}{2}+1)}$ 分别是位次为 $\frac{n+1}{2}$ 、 $\frac{n}{2}$ 、 $\frac{n}{2} + 1$ 对应的观察值。

②频数表法：用于连续性资料，可以通过百分位数法求 M 。

$$M = P_{50} = L + \frac{i}{f}(n \cdot 50\% - F_L)$$

L 为第 M 所在组段的下限值； i 为 M 所在组段的组距； f 为 M 所在组段的频数； F_L 为截止到 M 所在组段前的累计频数。

【例 6-2】 计算表 6-1 中的中位数 M 。

表 6-1 199 名食物中毒患者潜伏期

潜伏期(h) (1)	人数 f(2)	累计频数 Σf(3)	累计频率(%) (4) = (3)/n
0 ~	30	30	15.1
12 ~	71	101	50.8
24 ~	49	150	75.4
36 ~	28	178	89.4
48 ~	14	192	96.5
60 ~	6	198	99.5
72 ~84	1	199	100.0
合计	199		

用频数表法求 M 的具体计算步骤如下。

①计算 $\frac{n}{2}$ 大小，并按组段由小到大计算累计频数和累计频率，如表 6-1 第(3)、(4)栏。

②确定 M 所在组段。累计频数中大于 $\frac{n}{2}$ 的最小数值所在的组段为 M 所在的组段；或累计频率中大于 50% 的最小频率所在的组段为 M 所在的组段。

③按公式 $P_{50} = L + \frac{i}{f}(n \cdot 50\% - F_L)$ 求中位数 M 。

【实例解析】 本例 $n = 199$ ，根据表 6-1 第(2)栏数据，自上而下计算累计频数及累计频率，见第(3)、(4)栏。 $\frac{n}{2} = 99.5$ ，由第(3)栏知，101 是累计频数中大于 99.5 的最小值，或由第(4)栏知 50.8% 是大于 50% 的最小的累计频率，故 M 在“12 ~”组段内，将相应的 L 、 i 、 f 、 F_L 代入公式，求得 M 。

$$M = P50 = P_x = L + \frac{i}{f}(n \cdot x\% - F_L) = 12 + 12/71(199 \times 50\% - 30) = 23.75h(\text{小时})$$

4) 众数

众数：性质相同定量数据中出现次数(频数)最多的那个数为众数。

2. 离散程度指标描述定量资料

描述定量变量资料频数分布的另一主要特征是离散程度，用离散程度指标或变异指标表示。只有把集中指标和离散指标结合起来才能全面反映资料的分布特征。

离散程度指标种类：常用的描述定量资料离散程度的指标有极差、四分位数间距、方差、标准差、变异系数。

1) 极差(range)

极差即最大值减去最小值，也称全距，是一组同质观察值中最大值与最小值之差。它反映了个体差异的范围，全距大，说明变异度大；反之，全距小，说明变异度小，

(1) 计算： $R = x_{\max} - x_{\min}$ 。

(2) 优点：简便。缺点：信息利用不充分，稳健性较差。

用全距描述定量资料的变异度大小，虽然计算简单，但有诸多不足之处：只考虑最大值与最小值之差异，不能反映组内其他观察值的变异度；样本含量越大，抽到较大或较小观察值的可能性越大，则全距可能越大。因此样本含量相差悬殊时不宜用全距比较。

2) 四分位数间距 (quartile range, Q)

(1) 百分位数 (percentile, P_x)：它表示一组观察值按升序排列、等分为 100 等份、位居第 $x\%$ 位置的数，计算公式为 $P_x = L + \frac{i}{f}(n \cdot x\% - F_L)$ 。四分位数间距为上四分位数 Q_3 (即 P_{75}) 与下四分位数 Q_1 (即 P_{25}) 之差，即 $Q = P_{75} - P_{25} = Q_3 - Q_1$ 。四分位数间距可看成中间 50% 观察值的极差，其数值越大，变异度越大，反之，变异度越小。

(2) 适用资料：可以用于各种资料，特别是偏态分布资料。

(3) 由于四分位数间距不受两端个别极大值或极小值的影响，因而四分位数间距比全距稳定，但仍未考虑全部观察值的变异度，常用于描述偏态频数分布以及分布的一端或两端无确切数值资料的离散程度。

3) 方差 (variance)

每一个观察值与均数之差平方之和的平均值。为了全面考虑观察值的变异情况，克服全距和四分位数间距的缺点，需计算总体中每个观察值 X 与总体均数 μ 的差值 $(X - \mu)$ ，称为离均差。由于 $\Sigma(X - \mu) = 0$ ，不能反映变异度的大小，而用离均差平方和 $\Sigma(X - \mu)^2$ 反映之，同时还应考虑观察值个数 N 的影响。

(1) 方差计算：总体方差计算公式为 $\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$ ，实际应用中，总体均数往往是未知的。常用样本均数代替总体均数，样本方差估计总体方差。用样本例数 n 代替 N ，但用上式计算的结果总是比实际 σ^2 小。英国统计学家 W. S. Gosset 提出用 $n - 1$ 代替 n 来校正，这就是样本方差 s^2 ，其公式为：
$$S^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} = \frac{\Sigma x^2 - (\Sigma x)^2/n}{n - 1}$$
，式中的 $n - 1$ 称为自由度 (degree of freedom, df)。

若各观察值离均数越近，方差就越小，表示观察值离散小，数据较集中。各观察值离均数越远，方差就越大，表示各观察值离散度大，数据分布分散。

(2) 适用范围：方差适用于对称分布资料，尤其是正态或近似正态分布资料。

4) 标准差 (standard deviation, SD)

由于方差的度量单位是原度量单位的平方，为了方便实际使用，将方差公式开方，即得到标准差。总体标准差 $\sigma = \sqrt{\Sigma(X - \mu)^2/N}$ ，样本标准差 $S = \sqrt{\Sigma(X - \bar{X})^2/(n - 1)}$ 。

离均差平方和 $\Sigma(X - \bar{X})^2$ 常用 SS 或 l_{xx} 表示。数学上可以证明 $SS = l_{xx} = \Sigma(X - \bar{X})^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$ ，所以，样本标准差的计算公式可写成：

直接法：
$$S = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n - 1}}$$

加权法：
$$S = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{\sum f}}{\sum f - 1}}$$

(1)标准差的用途：反映个体观察值的离散程度大小；结合均数描述正态分布特征；计算其他指标，如标准误等。

(2)适用范围：标准差与方差适用范围相同，适用于对称分布资料，尤其是正态或近似正态分布资料。

5) 变异系数 (coefficient of variation, CV)

变异系数是标准差与均数之比，一般用百分数表示。

(1)计算：
$$CV = \frac{s}{\bar{x}} \times 100\%$$

(2)用途：可以用于比较度量衡单位不同资料的变异度；比较均数相差悬殊资料的变异度。CV 没有量纲(即单位)，消除了量纲的影响，CV 越大意味着相对于均数而言，变异程度越大。

归纳小结：定量资料常用的集中趋势和离散程度统计描述指标及适用资料类型归纳如表 6-2 所示。

表 6-2 定量资料统计描述常用的统计指标及适用资料类型

描述指标	指 标	意 义	适用资料类型
集中趋势指标	算术均数	个体的平均值	对称分布
	几何均数	平均倍数	取对数后对称分布
	中位数	位次居中的观察值	(1)非对称分布；(2)半定量资料；(3)末端无确切数值；(4)分布不明
	众数	频率最高的观察值	不拘分布，概略分析
	调和均数	基于倒数变换的平均值	正偏态分布资料
离散程度指标	全距	观察值取值范围	不拘分布，概略分析
	标准差(方差)	观察值平均离开均数的程度	对称分布
	四分位间距	居上半数观察值的全距	(1)非对称分布；(2)半定量资料；(3)两端无确切数值；(4)分布不明
	变异系数	标准差与均数的比值	(1)不同量纲的变量间比较；(2)量纲相同但数量级相差悬殊的变量间比较

3. 样例展示及辨析与释疑

【例 6-3】 测得某地 299 名正常人某项医学指标，频数表如表 6-3 所示，试用合适统计指标描述该医学指标的集中趋势和离散程度。

遇到频数表资料，在选择描述集中趋势和离散程度指标时，很多初学者很容易选择 $\bar{x} \pm s$ 描述，但是忽略了每个指标的适用条件，本例通过频数分布可以看出该指标分布为偏态分布，所以宜用中位数 M 来描述该医学指标的集中趋势，用四分位间距 Q 描述其离散程度。可通过以下公式计算求的中位数 M ，

$$M = P_{50} = L + \frac{i}{f}(n \cdot 50\% - F_L)$$

可通过以下公式计算求的中位数四分位间距 Q ，

$$Q = P_{75} - P_{25} \\ = \left[L + \frac{i}{f}(n \cdot 75\% - F_L) \right] - \left[L + \frac{i}{f}(n \cdot 25\% - F_L) \right]$$

表 6-3 某地 299 名正常人的某医学指标 (mg/L) 频数表

指 标	值 频 数
0 ~	31
8 ~	108
16 ~	67
24 ~	25
32 ~	13
40 ~	5
48 ~	3
56 ~ 64	4

6.1.3 如何合理应用统计描述指标描述定性资料

定性资料又称分类变量资料，是先将观察对象按某种性质分组，从而得到各组例数(频数)，这种资料为定性资料。描述定性资料指标通常用绝对数和相对数去描述。当我们用绝对数来表达资料时，往往因比较的基础相差悬殊而没有说服力；此时，就需要借助于相对数。

1. 相对数的种类

相对数(relative number)：是两个有联系的指标之比，是描述定性资料的主要统计指标。以相对数描述分类变量的水平可以消除由于基数不同导致的资料间的不可比。例如：甲、乙两地某病的发生例数分别为 100 例和 150 例，仅从发病例数难以比较两地的发病严重程度，如果已知两地的人口数均为 200000 人，则两地的发病率分别为 5‰和 7.5‰，很容易地判断出乙地的发病情况比甲地严重。

常用相对数类型包括频率(frequency)、强度(intensity)、相对比(relative ratio)。

(1)频率型指标：通常近似地反映某一事件出现的机会大小，表示各组频率在总例数中所占的比重，即反映内部各部分组成比例的一个指标，又称构成比，常以百分数表示，如发病概率、死亡概率等。

$$\text{频率} = \frac{\text{某一部分的观察数}}{\text{各部分的观察总和}} \times K$$

(2)强度型指标：指单位时间内某现象发生的频率，实质上为频率强度，用来反映事物的严重程度，实际运用中常称为率(rate)。常用的率包括发病率、患病率、死亡率、病死率等，应用计算各率时的不同，要特别注意不同的率之间意义不同，计算过程中率的分子、分母也不同，如发病率与患病率、死亡率与病死率。

$$\text{强度型指标} = \frac{\text{某事件发生的个体数}}{\sum (\text{可能发生某事件的个体数} \times \text{时间})} \times K$$

需要强调的一点是，在实际应用过程中构成比和率有时较难分清，人们在使用中经常混淆。因为两者的公式很相似，只是强度性指标(率)加入了时间的概念，指的是单位时间内发生的频率。它们的不同点是：率反映某种事物或现象发生的强度，而构成比则反映“部分与整体”或“某一部分与另一部分”之间的关系。上面公式中的 K 为比例常数，通常取 100% 或 1000‰。

(3)相对比型指标：相对比型指标是指任何两个相关联的变量 A 与 B 之比。实际应用中简称比(ratio)。例如：性别比、每千人口的医生数、CV 等。

$$\text{相比} = \frac{A}{B} \times 100\%$$

相互比较的两个指标可以是相同性质的指标，也可以是性质不同的指标；两变量可以为数值变量、分类变量，可以是绝对数、相对数、平均数等。常用的相对比有变异系数(CV)、流行病学中的相对危险度(RR)、性别比例、每千人口中医生数、每平方公里人口数、血清中白蛋白与球蛋白之比(A/G)等。

常用的相对比指标有三种。

- ①对比指标：指两个同类事物某指标的比，如男女性别比 106.04:100。
- ②关系指标：指两个有关的、但非同类事物的数量的比，如医护人员与病床数比。
- ③计划完成指标：指计划完成的程度，常用实际数达到计划数的%或几倍表示。

2. 应用相对数应注意的问题

(1)计算相对数的分母不宜太小。观察例数过小时，抽样误差较大，计算的相对数往往不稳

定,可靠性差。所以当观察例数较少(如少于30例)时,一般以绝对数表示为好,如以相对数表示,应给出其可信区间。但在动物实验中,由于实验过程中可以严格控制各项因素的影响,如严格挑选实验动物,控制实验条件,如温度、湿度等,最大限度地降低这些因素对实验结果的影响,这样使用较少的动物也可得到较稳定的实验结果,此时可以以相对数表示。

(2)不能用构成比代替率。构成比与率是两个不同的概念,其意义也不同,前者反映的是事物内部中各组成部分所占的比例,不能反映某事件发生的频率和频度,常见的错误为:根据构成比来比较不同事件的发生频度,或比较不同年代某事件的发生频率。

(3)正确计算平均率。求几个率的平均率时,容易简单地将几个率相加后除以其个数。相对数的分母(观察单位总数)不同,求平均数时的权数也应不同,因此应对各个率乘以各自的权数,或将各率的分子、分母分别相加后,分子之和除以分母之和求得平均率。

(4)相对数比较时应具有可比性。忽视资料的可比性,将各相对数直接相比较。影响相对数大小的因素很多,除研究因素外,其他因素,如性别、年龄、病情等也可影响相对数的大小。因此,应注意消除其他因素的影响,使相互比较的相对数具有可比性。

(5)样本率或样本构成比比较时应做假设检验。如数值型变量一样,抽样所得的样本相对数也有抽样误差,因此相对数间的比较也应做假设检验。而不应该将两个样本间的率直接进行比较、简单下结论。

3. 样例展示及辨析与释疑

【例6-4】 表6-4中数据为某地1993—1998年6年间意外损伤、中毒发生情况和死亡情况。有人根据下表得出结论:1998年病死率最高。

表6-4 某地1993—1998年6年间意外损伤、中毒发生情况和死亡情况

年度(1)	发病例数(2)	死亡例数(3)	病死率(%) (4)	死亡构成比(%) (5)	发病率相对比(6)
1993	584	8	1.37	8.8	—
1994	571	10	1.75	11.0	1.28
1995	714	12	1.68	13.2	1.23
1996	748	16	2.14	17.6	1.56
1997	942	21	2.23	23.0	1.63
1998	1095	24	2.19	26.4	1.60
合计	4654	91	1.96	100.0	—

【实例解析】 表6-4中第(2)栏、第(3)栏分别为各年的发病例数和死亡例数,第(4)栏为病死率,即第(3)栏与第(2)栏的比值,如1993年病死率为 $8/584=1.37\%$;第(5)栏为死亡构成比,即各年度死亡数占总死亡人数的百分比,是第(3)栏与总死亡人数的比值,如1993年的死亡构成比为 $8/91=8.8\%$;相对比为各年的发病率与1993年发病率之比,如1994年与1993年相比,前者是后者的 $1.75/1.37=1.28$ 倍。

本例计算过程既包含率、构成比,也包括相对比的运算。在实际应用过程中,率和构成比两者非常容易混淆,我们在应用过程中应该紧密结合各个相对数的概念进行区分和分析运算。

【例6-5】 为了检验一种少数民族的中药偏方治疗某种罕见疾病的治疗效果,随机抽取了8例罕见疾病患者,用该中药偏方进行治疗,经过一段时间后发现4例痊愈,而4例未愈,所以下结论得出这种少数民族的中药偏方治疗这种罕见疾病的治疗有效率为50%。

【实例解析】 本实例中计算相对数的分母太小,观察例数过少时,抽样误差较大,计算的相对数往往不稳定,可靠性差。缺乏代表性会造成相对数不稳定,不能准确地反映总体的客观规律。例如,在药物临床试验中,如果观察例数是2,那么,结果可能是2例都有效,可能是2例都无效,也

可能是 1 例有效另 1 例无效。从统计学角度看, 由于观察单位数太少, 任何偶然的因素都会造成结果的不稳定, 因此无论哪种结果发生, 都无法确信结果的可靠性。在临床试验研究中, 如果观察例数确实很少, 则应当用绝对数来表示结果, 或者用后面将要介绍的统计推断知识, 计算出率的可信区间; 更科学、严谨的做法是: 当例数很少时, 不值得开展临床试验研究, 即使有了试验数据, 也不值得进行统计分析。

【例 6-6】 某研究者检查专业人员的矽肺的患病情况, 得到表 6-5 中的数据。通过表 6-5 第三列发现, 在 147 名患者中, 10~15 专业工龄段的患病人数最多, 而且第四列也显示出专业工龄为 10~15 年的工人所占的比例最大(占 49.66%), 明显高于其他专业工龄段, 因此下结论: 认为专业工龄为 10~15 年的患病率最高。

【实例解析】 本实例犯了典型的用构成比代替率的错误。构成比表示事物内部各部分所占的比例, 不能用构成比来说明某现象发生的频率或强度。事实上, 因为专业工龄为 10~15 年的工人最多, 所以患病的人数就可能多。至于哪一组工人的患病率高, 应当比较各组的患病率, 第三列分别除以第二列得到各组段的患病率, 正确分析结果见表 6-6, 从表 6-6 的最后一列可以看出, 最后一个组的患病率最高, 即工龄越长, 患病率就越高。

表 6-5 某研究者检查专业人员的矽肺的患病情况

专业工龄(年)	检查人数	患病人数	构成比(%)
1~	340	17	11.56
5~	254	30	20.41
10~	432	73	49.66
15~	136	27	18.37
合计	1162	147	100.00

表 6-6 某研究者检查专业人员的矽肺的患病情况

专业工龄(年)	检查人数	患病人数	构成比(%)	患病率(%)
1~	340	17	11.56	5.00
5~	254	30	20.41	11.81
10~	432	73	49.66	16.90
15~	136	27	18.37	19.85
合计	1162	147	100.00	12.65

【例 6-7】 某研究者拟比较西莫地平和尼莫地平两种药物治疗高血压的有效率, 搜索以往权威杂志, 发现有一篇文献报道西莫地平治疗高血压病人 100 例, 治疗有效人数为 60 人; 一篇文献报道尼莫地平两种药物治疗高血压病人 100 例, 治疗有效人数为 85 人; 该研究者由此得出尼莫地平治疗高血压的有效率高于西莫地平。

【实例解析】 研究者由文献报道西莫地平治疗高血压的有效率为 60%, 另一篇文献报道尼莫地平治疗高血压的有效率为 85%, 由此得出尼莫地平治疗高血压的有效率高于西莫地平的结论。其实, 研究者犯了以下两个错误。

(1) 没有考虑到相对数比较时应具有可比性; 两篇文献报道可能调查地区, 或者调查人群的性别比或者年龄构成, 病情严重程度等不同, 这样直接比较无法消除这些重要非试验因素(它们在两组之间不平衡, 就变成了混杂因素) 的干扰, 容易得出误差较大甚至是错误的结论。必要时需要进行率的标准化的, 才能进行比较。

(2) 已知条件只能获得两个调查的样本有效率, 样本有效率存在抽样误差, 所以不能直接比较, 也不具备进行假设检验的前提条件。

本部分实际应用过程中的错误主要体现在前面“应用相对数应注意的问题”中叙述的几个方面，应用者在实际运用过程中一定要加以注意。

6.2 如何合理应用统计图、统计表

统计表和统计图是统计描述的重要方法。医学科学研究资料经过整理和计算各种统计指标后，所得结果除了用适当的文字说明外，常将统计资料及其指标以表格列出(称为统计表，即 statistical table)，或将统计资料形象化，利用点的位置、线断的升降、直条的长短或面积的大小等形式直观表示事物间的数量关系(称为统计图，即 statistical graph)。统计表与统计图可以代替冗长的文字叙述，表达清楚，对比鲜明。

6.2.1 统计表

统计表(statistical table)：在研究报告和科研论文中，常将统计分析指标及其结果以表格的形式列出，称为统计表；是描述资料特征，呈现统计分析结果的主要工具。

1. 统计表结构组成

统计表由标题、标目、线条、数字四部分组成，有时还需要备注部分。

2. 统计表结构种类

(1)简单表(simple table)由一组横标目和一组纵标目组成，如表 6-7 所示。

表 6-7 2001 年某省不同地区的卫生系统反应性评分比较

地 区	调 查 人 数	评 分 均 值
省会城市	333	703.63
一般城市	152	507.15
县及乡村	971	679.06
合计	1456	666.73

(2)复合表(combivative table)是由两组及以上的横标目和纵标目相结合起来或一组横标目和两组及以上纵标目结合起来以表达它们之间关系的统计表，如表 6-8 所示。

表 6-8 2001 年某省不同地区、性别的卫生系统反应性评分比较

地 区	调 查 人 数		评 分 均 值	
	男	女	男	女
省会城市	217	116	706.60	698.07
一般城市	100	52	517.15	487.92
县及乡村	371	600	669.88	684.74
合计	688	768	659.26	673.43

3. 统计表绘制要求

(1)简明扼要，重点突出，一张表格最好只表达一个中心内容。

(2)应主谓分明，层次清楚，合理安排主语和谓语的位置，统计表编制完成后，从左向右读依次横标目(主语)和纵标目(谓语)，应构成完整的一句话。

(3)要认真核对表内数据，以保证准确可靠。

4. 统计表编制要点

(1)标题：写在表的上方中央；简明扼要地概括出表的主要内容；必要时应注明资料收集的时间、地点。应用过程中常见错误有：过于简单；过于烦琐；题意不确切、具体等。

(2)标目：标目是表格内的项目；根据位置和作用可以分为横标目、纵标目和总标目；文字应简明，有度量单位时应注明。

横标目位于表最左侧，说明横行的数字含义，是表的主语位置，是被研究的对象。

纵标目位于表的右上方，说明纵列数字的含义，是表的谓语位置，标志被研究对象的各项指标。

应用过程中常见错误有：标目层次过多；横标目和纵标目顺序颠倒；含意不确切；有单位的标目没有注明单位等。

(3)线条：统计表常仅有三条基本线，即顶线、底线及表头与表身之间的分隔线三条线，若有第二个分组标志，还需要在表头中间加一条短横向辅助线，将总的纵标目与第二个分组标志分隔开。当有合计栏时，还有一条合计栏线。但表格中不应有竖线，斜线。

(4)数字：准确，用阿拉伯数字表示。同一指标小数点位数应一致，位置应对齐。表格内不留空格，无数字时用“—”表示；缺失或未记录数字时用“…”表示；数字若为“0”，应如实填写“0”。

(5)备注：不是表格的主要研究内容，但又需要在列表时表达的内容称为备注。备注不列在表格内，如表内有需要注明的内容，先用“*”在表内做出标记，把需要表达的内容在底线下加以说明。

5. 统计表样例展示及辨析与释疑

【例 6-8】 表 6-9 表示出两组患者健康教育满意度及健康知识评分比较。试分析有无不妥之处。

表 6-9 两组患者健康教育满意度及健康知识评分比较 ($\bar{x} \pm s$, 分)

组 别	例 数	满意度评分	健康知识评分
对照组	50	95.11 \pm 3.01	86.82 \pm 3.28
研究组	50	99.42 \pm 2.76	93.47 \pm 1.80
<i>t</i>		22.51	12.69
<i>P</i>		<0.05	<0.05

【实例解析】 表 6-9 存在一个明显不足：在表格的最后两行出现统计量 *t* 值与 *P* 值，与总横标目“组别”是不符的，因此，这样的安排并不合理，应删除最后两行，统计分析结果可以在论文中体现。修改后的统计表略。

【例 6-9】 在某文献中，研究者采用艾森克人格问卷 (EPQ) 测试干预组与对照组官兵的人格维度，得分见表 6-10。

表 6-10 修改前的干预组与对照组 EPQ 各人格维度得分 ($\bar{x} \pm s$, 分)

人 格 维 度	对 照 组	干 预 组	<i>P</i> 值
神经质	6.48 \pm 3.43	8.13 \pm 4.95	<0.01
内外向	12.98 \pm 3.79	11.31 \pm 5.34	<0.01
精神质	11.74 \pm 5.17	9.63 \pm 5.21	<0.01
掩饰倾向	11.48 \pm 3.95	11.12 \pm 4.24	>0.05

【实例解析】 表 6-10 不符合规范。主要是主、谓语倒置，每张表格都有主语和谓语，主语指被研究的事物，谓语说明主语的各项指标；一般横标目代表主语，纵标目代表谓语，两者结合起来表达一个完整的意思。当主语比较简单、谓语较多时，可以用纵标目代表主语，横标目代表谓语。修改后的正确表格如表 6-11 所示。

表 6-11 修改后的干预组与对照组 EPQ 各人格维度得分 ($\bar{x} \pm s$, 分)

试 验 分 组	EPQ 评分			
	*：神经质	内 外 向	精 神 质	掩 饰 倾 向
对照组	6.48 ± 3.43	12.98 ± 3.79	11.74 ± 5.17	11.48 ± 3.95
干预组	8.13 ± 4.95	11.31 ± 5.34	9.63 ± 5.21	11.12 ± 4.24

注：* 代表人格维度。

【例 6-10】 表 6-12 为 1592 例口腔颌面部肿瘤、囊肿和瘤样病变例数和百分比，试指出其中的错误。

表 6-12 1592 例口腔颌面部肿瘤、囊肿和瘤样病变例数和百分比

病 变 性 质			例 数			百分比 (%)	
肿瘤	良 性		674		1232	42. 34 35. 05	77. 39
	恶 性	癌	437	558			
		肉瘤	101				
		其他	20				
囊肿			192		12. 06		
瘤样病变			168		10. 55		
总计			1592		100. 00		

【实例解析】 本实例中存在的问题较多，首先线条过多，没有采用三线格形式，而且还存在竖线；其次统计表层次不清，结构混乱，难于理解。按照统计表的编制原则和要点修改后的统计表如表 6-13 所示。

表 6-13 口腔颌面部不同病变构成情况

病 变 性 质	例 数	构成比 (%)
良性肿瘤	674	42.34
恶性肿瘤 *	558	35.05
囊肿	192	12.06
瘤样病变	168	10.55
合计	1592	100.00

注：* 代表包括癌 437 例、肉瘤 101 例、其他恶性肿瘤 20 例。

【例 6-11】 表 6-14 是不同溶液冲洗伤口的效果比较，试找出其中的错误。

表 6-14 不同溶液冲洗伤口的效果比较

	试 验 组			对 照 组
	庆 大 霉 素	新 洁 尔 灭	四 环 素	生 理 盐 水
总例数	30	30	30	30
感染例数	1	3	5	8
百分比	3.3%	10%	16.7%	26.7%

【实例解析】 该实例存在的主要问题有横标目和纵标目倒置；比例基数未放置在标目中；多余的横竖线条；误将感染率写成了百分比。修改后的表格如表 6-15 所示。

表 6-15 修改后的不同溶液冲洗伤口效果比较

溶 液	总 例 数	感 染 例 数	感染率(%)
庆大霉素	30	1	3.3
新洁尔灭	30	3	10.0
四环素	30	5	16.7
生理盐水	30	8	26.7

【例 6-12】 表 6-16 是不同体重初生儿的死亡率。

表 6-16 不同体重初生儿的死亡率

总 活 产 数	5674			
体重分组	<1500g	~2000g	~2500g	≥2500g
新生儿数	14	61	280	5292
死亡数	11	10	12	24
死亡率	785.7‰	163.9‰	42.9‰	4.5‰

【实例解析】 该实例存在的主要问题有标目设置不当，横纵标目未分，主谓不明；合计项设置不当；体重分组表示不规范；比例基数未放置在标目中等。修改后的表格如表 6-17 所示。

表 6-17 某医院 1980—1983 年不同体重初生儿的死亡率

初生儿体重(g)	新 生 儿 数	死 亡 数	死亡率(‰)
<1500	14	11	785.7
1500~	61	10	163.9
2000~	280	12	42.9
≥2500	5292	24	4.5
合计	5647	57	10.1

6.2.2 统计图

统计图(statistical chart)：是用点的位置、线段的升降、直条的长短、面积的大小等来表达统计数据的一种统计表达工具。

1. 统计图绘制要求

- (1)按资料的性质和分析目的选用最合适的图形。
连续性资料一般常用线图、直方图描述，离散型资料一般常用条图、圆图、百分条图描述。
- (2)要有标题：概括图形所要表达的主要内容，标题应包括研究资料的时间、地点，一般位于图形的下端中央。
- (3)条图、直方图、散点图、线图都有纵、横坐标轴，并表明尺度，数量从小到大等距标明。条图、直方图纵坐标必须从“0”开始。纵横轴的比例一般为 5:7。
- (4)对于不同事物或对象，宜选用不同线条和颜色表达，并附图例加以说明。图例一般位于右上角空隙处或图的下方。

2. 统计图的种类和适用资料

常见的统计图包括条图、圆图、百分条图、线图、半对数线图、直方图、散点图、统计地图等图形。

1) 条图 (bar graph)

条图用等宽长条的高度表示按性质分类资料各类别的数值大小,用于表示它们之间的对比关系,一般有单式与复式之分。

条图制图要求:

- (1) 一般以横轴表示分组变量的水平,表示各个类别;纵轴表示其数值大小。
- (2) 纵轴尺度必须从“0”开始,中间不宜折断,否则会改变各对比组之间的比例关系;在同一图内,尺度单位代表同一数量时,必须相等。
- (3) 各直条宽度应相等,各直条之间的间隙也应相等,其宽度与直条的宽度相等或为直条宽度的 1/2。
- (4) 直条的排列通常由高到低或由低到高,以便比较。
- (5) 复式条图的绘制方法同上,所不同的是,复式条图有两个或两个以上的分组变量,外层分组变量的各水平包括内层分组变量的全部水平,直条所表示的类别应用图例说明,内层分组变量的各直条间不留空隙。图 6-1 为复式条图。

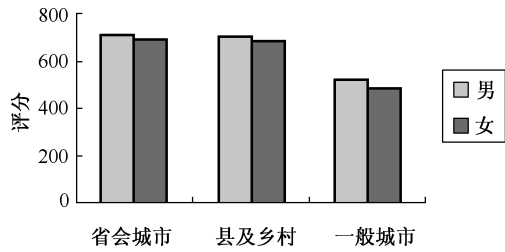


图 6-1 2001 年某省不同地区、性别的卫生系统反应性评分比较

2) 圆图 (pie graph)

圆形图适用于百分构成比资料,表示事物各组成部分所占的比重。以圆形的总面积代表 100%,把面积按比例分成若干部分,以圆心角的角度大小来表示各部分所占的比重。图 6-2 是 2001 年某医科大学专业技术人员构成的圆图。

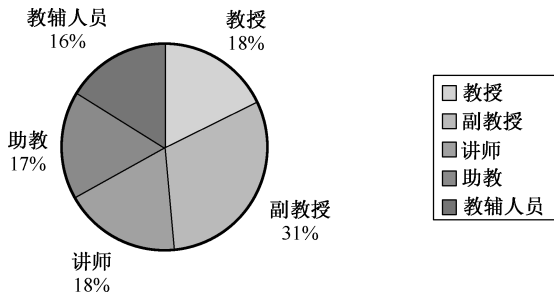


图 6-2 2001 年某医科大学专业技术人员构成

圆图制图要求:

- (1) 先绘制大小适当的圆形。由于圆心角为 360°, 因此每 1% 相当于 3.6° 的圆心角, 将各部分百分比分别乘以 3.6° 即各构成部分应占的圆心角度数。
- (2) 圆形图上各部分自圆的 12 点开始由大到小按顺时针方向依次绘制, 其他置最后。所得各部分的扇形面积即代表某一构成部分。
- (3) 圆中各部分用线分开, 注明简要文字及百分比或用图例。

(4) 如有 2 种或 2 种以上性质类似的资料相比较, 应绘直径相同的圆, 并使各圆的圆心落在同一条水平线上并使圆中各部分的排列次序一致, 以利于比较。

3) 百分条图 (percent bar)

百分条图的意义及适用资料与圆形图相同, 不同的是表现形式不一样。百分条图也称构成条图, 是以直条总长度作为 100%, 直条中各段表示事物各组成部分构成情况, 如图 6-3 所示。

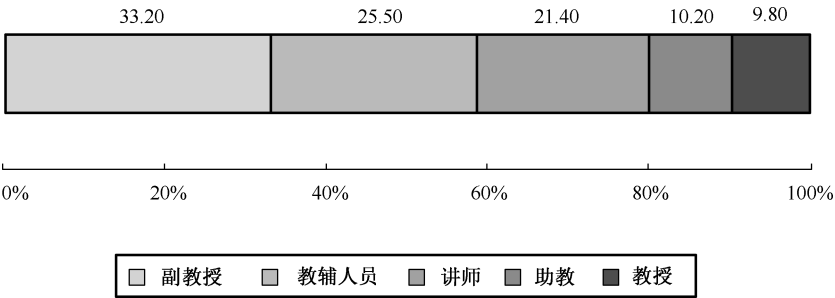


图 6-3 2001 年某医科大学专业技术人员构成

百分条图制图要求:

- (1) 先绘制一个标尺, 尺度分成 5 格或 10 格, 每格代表 20% 或 10%, 总长度为 100%, 尺度可绘制在图的上方或下方。
- (2) 绘一直条, 全长等于标尺的 100%, 直条宽度可任意选择, 一直条内相对面积的大小代表数量的百分比。
- (3) 直条各部分用线分开并注明简要文字及百分比或用图例表示。
- (4) 资料一般按各构成由大到小, 自左至右依次排列, 其他置后。

4) 线图 (line graph)

线图适用于连续性资料, 以不同的线段升降来表示资料的变化, 并可表明一事物随另一事物 (如时间) 而变动的情况。常见的有纵横轴均为算术尺度, 表示时间变化趋势的普通线图。如果纵轴为对数尺度, 横轴为算术尺度, 表示消长趋势的半对数线图 (semilogarithmic line graph)。图 6-4 是某地 1950—1966 年伤寒与结核病死亡率随时间变化的线图。

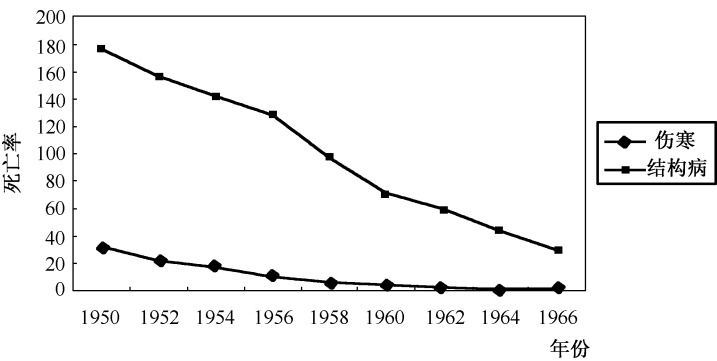


图 6-4 某地 1950—1966 年伤寒与结核病死亡率随时间变化的线图

普通线图 (line chart): 用线段的升降来表示一事物随另一事物变化的趋势, 适用于表达一组连续性资料 (如时间、年龄等) 的变化“幅度”。

半对数线图 (semi-logarithmic linear chart): 是线图的一种特殊形式, 其纵轴为对数尺度,

横轴为算术尺度，使线图上的数量关系变为对数关系。特别适宜于表达多组相同指标变化“速度”。

普通线图绘制要求：

- (1)横轴表示某一连续变量(时间或浓度或年龄等)；纵轴表示某种率或频数，其尺度必须等距(或具有规律性)。
- (2)同一图内不应有太多的曲线，通常 ≤ 5 条，以免观察不清。
- (3)如有几根线，可用不同的图线(实线、虚线等)来表示，并用图例说明。
- (4)图线应按实际数字绘制成折线，不能任意改为光滑曲线。

绘制半对数线图的方法可使用特制的半对数坐标纸，也可将统计指标取对数值后在方格坐标纸上作图。其具体绘制方法与线图相同。

5) 直方图 (histogram)

直方图用于表达连续性资料的频数分布；以不同直方形面积代表数量，各直方形面积与各组的数量成正比关系。图 6-5 为直方图。

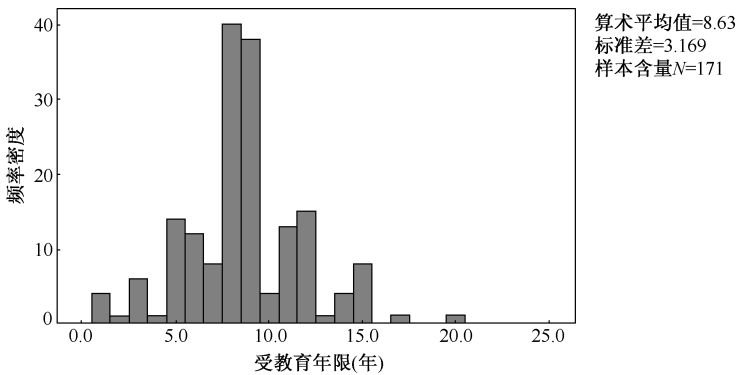


图 6-5 2001 年某地居民受教育年限频数分布

直方图制图要求：

- (1)一般纵轴表示被观察现象的频数(或频率)，横轴表示连续变量，以各矩形(宽为组距)的面积表示各组段频数。
- (2)纵轴的刻度必须从“0”开始，而横轴刻度只需标出实际范围。
- (3)直方图的各直条间不留空隙；各直条间可用直线分隔，但也可不用直线分隔。
- (4)各矩形的高度为频数或频率，宽度为组距。组距不等时，横轴仍表示连续变量，但纵轴是每个横轴单位的频数，则应调整各矩形高度：矩形高度 = 组段频数/组距。

6) 散点图 (scatter diagram)

散点图以直角坐标系中各点的密集程度和趋势来表示两现象间的关系。例如某研究者拟研究饮用水氟含量与氟骨症患病率两者的关系，根据收集的资料如表 6-18 所示。

表 6-18 某地区饮用水氟含量与氟骨症患病率

氟含量 (mg/L)	20	21	25	28	30	35	40	45
患病率 (%)	0.6	0.8	1.0	1.6	1.8	3	3.5	4.5

根据表 6-18 中的数据绘制两者关系散点图，如图 6-6 所示。根据点的散布情况，推测两种事物或现象有无相关，故常在对资料进行相关或回归分析之前使用。

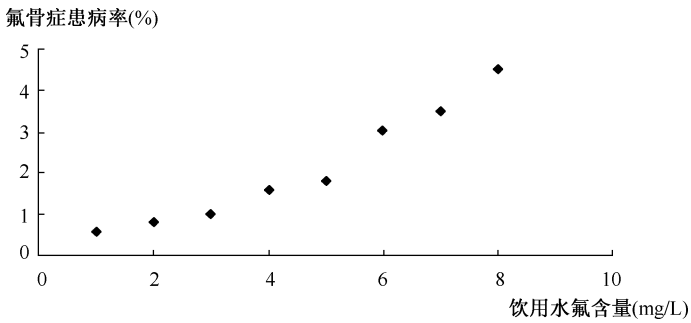


图 6-6 某地区饮水氟含量与氟骨症患病率散点图

- 散点图制图要求：
- (1)一般横轴代表自变量或可进行精确测量、严格控制的变量，纵轴则代表与自变量有依存关系的因变量。
 - (2)纵轴和横轴的尺度起点可根据需要设置。
- 常用统计图的适用资料类型及用途见表 6-19。

表 6-19 常用统计图的适用资料类型及用途

统计图形	资料类型	用 途
条图	分类变量	直线长短表达数值大小
百分条图/圆图	分类变量	表达各部分所占比例，整体为 100%
线图	连续型变量	表达事物的动态变化(绝对差值)
半对数线图	连续型变量	表达事物的发展速度(相对比)
直方图	连续型变量	用直条的高度或面积表达各组段的频率或频数
散点图	连续型双变量	表达两变量的相互或依赖关系
统计地图	分类变量	表达疾病相关指标在地域上的分布
箱式图	连续型变量	用“箱”和“触须”等的位置表示变量的分布特征

3. 统计图样例展示及辨析与释疑

【例 6-13】 在《老年阵发性心房颤动发作的危险因素及其护理对策》一文中，研究者总结了老年阵发性房颤患者发作的诱因，见表 6-20。问：有没有表达该资料的更恰当的方式？

表 6-20 老年阵发性房颤患者发作的诱因

项 目	感 冒	情绪激动	失 眠	疲 劳	饱餐、腹胀	吸 烟	心肌缺血	无 诱 因
发作次数(n)	56	41	34	28	20	15	7	15
百分率(%)	25.9	19.0	15.8	13.0	9.3	6.9	3.2	6.9

- 【实例解析】 若用百分条图或圆图来表达资料，则更为恰当。统计图可以给人直观、深刻的印象。
- 【例 6-14】 某研究者收集了结核病和白喉两种呼吸系统传染病近 9 年的历年发病率资料。这两种疾病在建国后随着我国公共卫生健全，都有明显下降趋势，研究者欲通过收集的历年数据比较哪一种疾病的下降趋势更明显，绘制了线图(如图 6-7 所示)。由图 6-7 可以看出结核病的下降趋势更明显。

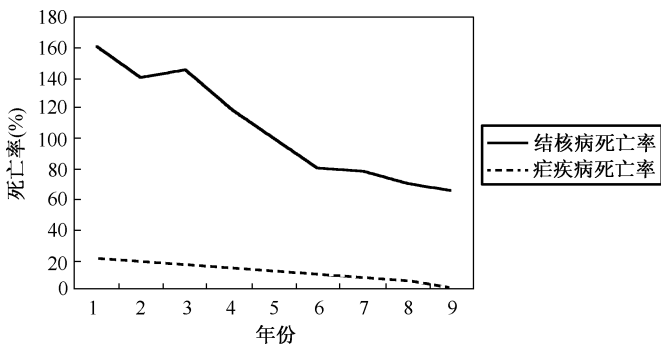


图 6-7 近 9 年结核病和白喉两种呼吸系统传染病的发病率 (普通线图)

【实例解析】 本实例为历年的发病率的数据，属于连续性资料随时间变化的趋势，适宜用线图表示。但是，研究者拟通过收集的历年数据比较哪一种疾病随时间下降的趋势更明显。这是涉及下降速度的问题。普通线图侧重表示一事物随另一事物变化的“幅度”；半对数线图特别适宜表达两组或多组相同指标的变化“速度”。所以本资料应该用半对数线图表达。绘制的半对数线图如图 6-8 所示，由此可以看出白喉的下降“速度”更明显。

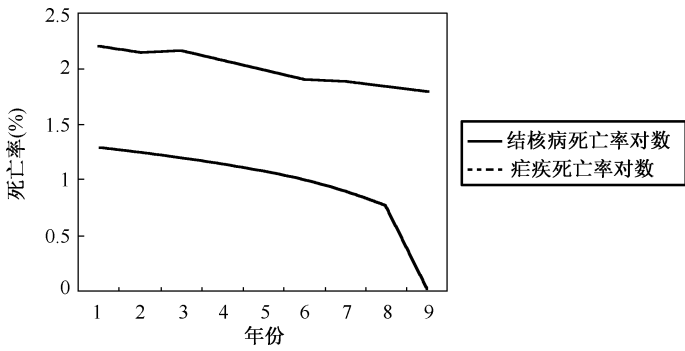


图 6-8 近 9 年结核病和白喉两种呼吸系统传染病的发病率 (半对数线图)

【例 6-15】 研究者欲了解三种肝炎在 A、B 地区两地的发病率(1、2、3 分别代表甲肝、乙肝、丙肝)，资料如表 6-21 所示。研究者对 A 地区和 B 地区肝炎发病率分别做出条形图(如图 6-9 和图 6-10 所示)，可直观地得出结论：三种肝炎的发病率均是 B 地区高于 A 地区。

表 6-21 三种肝炎在 A、B 两地区的发病率

肝 炎 类 型	发 病 率 (%)	
	A 地 区	B 地 区
甲 肝	0.2	0.15
乙 肝	0.8	0.32
丙 肝	0.02	0.13

【实例解析】 尽管通过这两幅图中各类型肝炎直条图的高低可以很直观地得出结论，三种肝炎发病率均是 B 地区高于 A 地区，但细心的读者会发现，两条形图软件自动给出的纵轴的单位长度不一致，B 地区高于 A 地区仅是错觉而已。可见，在用统计图表达同类数据时，纵轴上的单位长度要统一。

正确的画法是将两地区绘制在一个复式条形图中，既可以保证纵轴的单位长度统一，又可以方便地进行两两比较，如图 6-11 所示(注：左侧蓝色代表 A 地区，右侧红色代表 B 地区)。由此图可

以看出三种肝炎发病率只有丙肝是 B 地区高于 A 地区，其他两类肝炎发病率均是 A 地区高于 B 地区。

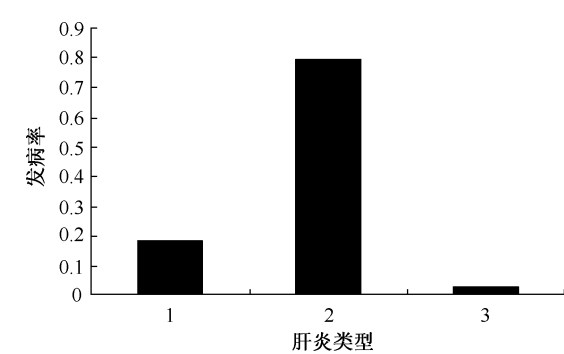


图 6-9 A 地区三种肝炎的发病率

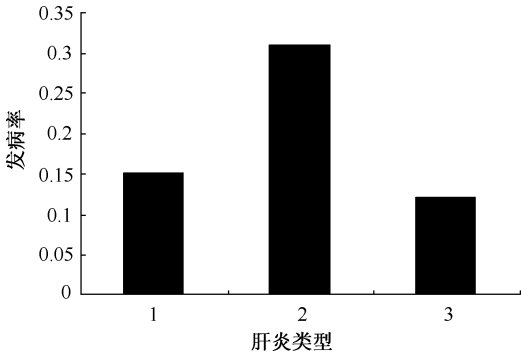


图 6-10 B 地区三种肝炎的发病率

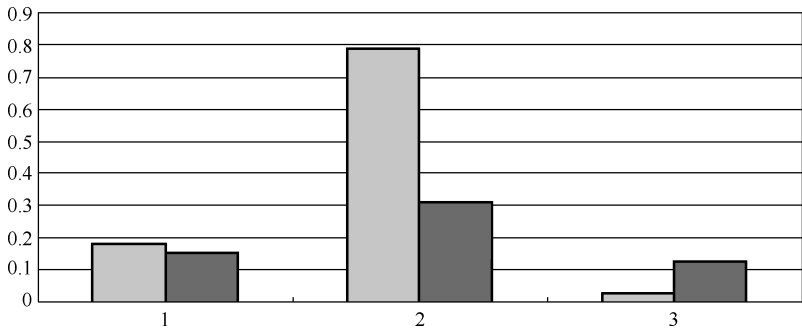


图 6-11 A 地区、B 地区三种肝炎的发病率比较

6.3 如何合理选用统计分析方法

在医学科学研究中，科研资料不仅可以选用合适的统计图、统计表及统计指标对其数量特征及分布规律进行测定和描述，而且在很多情况下，还需要通过合适的方法阐明事物间的内在联系和规律性，即对资料进行统计推断。在资料分析时，只有恰当、合理地选择统计推断指标和方法，方可真实地认识事物间的关系及其规律。当然，选择恰当、合理的统计方法对资料进行统计分析时，不仅需要考虑研究目的、资料类型、样本含量的大小、各种统计分析方法的应用条件等，还需要结合资料的具体科研设计方案和类型。

据了解，一些内科学类研究论著及杂志中发表的许多科研论文，存在错误应用统计分析方法处理资料的情况，多因素统计分析方法误用情况尤为严重。下面通过内科学杂志中的一些实例，按照资料类型就常用的一些统计分析方法的合理应用做一简单介绍。

6.3.1 定量资料统计分析方法的合理选用

选用统计分析方法分析统计资料时，需结合诸多因素，如需综合考虑设计方案、研究目的、资料类型等方面选择恰当的统计分析方法进行处理，但许多研究者在进行统计分析时，未进行综合考虑，只是根据其中的一个方面就错误、盲目地对资料进行统计分析。例如，在两个样本均值比较时，有些人不对资料进行分析，不管资料是否满足参数检验条件，也不管试验的设计类型如何，统

统以 $\bar{x} \pm s$ 的形式表示,并选用两独立样本 t 检验进行处理。另外,也有一些研究者对多因素设计试验随意进行“分割”,并采用单因素统计方法处理多因素设计资料。下面是笔者在研读内科学杂志中的学术论文时发现的有明显统计分析方法应用错误的一些实例,现将其列举出,并对其进行辨析和释疑。

1. 误用表达定量资料的简化方法表示定性资料

【例 6-16】 早期经鼻空肠管给予中药治疗重症急性胰腺炎的临床观察。原文统计学处理:应用 SPSS 10.0 软件处理,计量资料采用配对 t 检验,计数资料以 $\bar{x} \pm s$ 表示,采用 χ^2 检验。

【实例解析】 本例原文作者在选用统计方法时提及“计量资料采用配对 t 检验,计数资料以 $\bar{x} \pm s$ 表示,采用 χ^2 检验”。对于呈对称分布(最好呈正态分布)的定量资料,通常以 $\bar{x} \pm s$ 的形式表示,定性资料一般采用率或构成比等指标表示。故本例所犯错误是:误用定量资料统计描述指标来表示定性资料。事实上,原作者所说的“计数资料”在本质上应称为“定性资料”,不是像“脉搏次数/分钟”那样的真正的“计数资料”。

2. 所交代的统计方法与资料类型不相符

【例 6-17】 某研究者欲探讨炎症性肠病临床特点,以期提高对本病的认识。对某医院诊断为炎症性肠病的 390 例住院病例(其中溃疡性结肠炎 338 例,克罗恩病 52 例)进行回顾性研究。原文统计学处理:采用统计描述进行分析,数据处理均使用 SPSS 13.0 软件包。计数资料行 χ^2 检验,计量资料行 t 检验或 Fisher's 检验。 $P < 0.05$ 为差异有统计学意义。

【实例解析】 本例研究者在选用统计方法时提及“计数资料行 χ^2 检验,计量资料行 t 检验或 Fisher's 检验”。我们知道,在选择统计方法处理资料时,需综合考虑研究目的、试验设计类型、资料类型、样本含量大小等因素。 t 检验一般用于单样本或两样本定量资料均数间的比较(需满足参数检验条件),而“Fisher's 检验”应该指的是 Fisher's 精确概率法,它常用于对定性资料进行统计分析。而本例在文中提到“计量资料行 t 检验或 Fisher's 检验”。故本例研究者思路不清楚,犯了所选用的统计分析方法与资料类型不相符的错误。原作者所说的“计数资料”也不是真正的“计数资料”,而是“定性资料”。

3. 误用成组设计 t 检验处理配对设计一元定量资料

【例 6-18】 某研究者欲研究Ⅲ度房室传导阻滞患者高位右室间隔部起搏(HRVS)时 VAT 与 DDD 模式对心功能的影响。对 32 例Ⅲ度房室传导阻滞患者,在 HRVS 时,分别给予 DDD 模式或 VAT 模式工作,于调控即刻通过超声心动图测定二尖瓣口快速充盈期峰值血流速度(E_p)、二尖瓣口左房收缩期峰值血流速度(A_p)、二尖瓣环后壁处收缩期脉冲组织多普勒峰值速度平均值(V_s)、二尖瓣环后壁处舒张早期脉冲组织多普勒峰值速度平均值(V_e)、二尖瓣环后壁处舒张晚期脉冲组织多普勒峰值速度平均值(V_a)等指标。两种模式下各指标见表 6-22。统计分析:计量资料以 $\bar{x} \pm s$ 表示,采用两样本 t 检验进行处理。

表 6-22 二尖瓣口血流频谱与 DTI 参数比较 ($\bar{x} \pm s$, cm/s, $n=32$)

类 别	E_p	A_p	V_s	V_e	V_a
DDD	79.12 ± 9.06	81.92 ± 11.10	12.33 ± 1.08	13.16 ± 1.05	14.97 ± 1.57
VAT	80.28 ± 8.27	80.07 ± 10.68	12.26 ± 0.79	12.86 ± 1.10	13.31 ± 1.47

【实例解析】 该资料的观测指标为二尖瓣口快速充盈期峰值血流速度(E_p)、二尖瓣口左房收缩期峰值血流速度(A_p)、二尖瓣环后壁处收缩期脉冲组织多普勒峰值速度平均值(V_s)、二尖瓣环后壁处舒张早期脉冲组织多普勒峰值速度平均值(V_e)、二尖瓣环后壁处舒张晚期脉冲组织多普勒

峰值速度平均值(V_a),五个指标均为定量的观测指标。本资料研究者对 32 例Ⅲ度房室传导阻滞患者,在 HRVS 时,分别给予 DDD 模式或 VAT 模式工作,然后测定各指标。由此可知,每例患者均接受了两种不同的模式,此资料应为同一研究对象分别接受两种处理,然后比较一种处理因素两个水平下效果(指标)之间的差别大小,故属于配对设计定量资料。但原文作者在统计分析时提及“计量资料以 $\bar{x} \pm s$ 表示,采用两样本 t 检验进行处理”。由此可知,原文作者误将配对设计定量资料作为成组设计定量资料,用成组设计一元定量资料 t 检验处理配对设计一元定量资料,这种做法是错误的。假设本例资料各指标满足参数检验条件,则应选择配对设计一元定量资料 t 检验对资料进行处理。

另外,如果要同时考察两种不同工作模式下各指标的变化情况,本资料应属于两因素设计。由表 6-22 可知,本例的效应指标是二尖瓣口快速充盈期、左房收缩期、环后壁处收缩期、环后壁处舒张早期和环后壁处舒张晚期各期的血流速度,故二尖瓣分期是一个试验因素(时间与部位的复合因素)。另外,DDD 模式或 VAT 模式是两种不同的工作模式,所以工作模式是另一个试验因素。所以该资料可看成两因素的某种设计类型下的定量资料。由于两个因素都作用于每位患者,故该定量资料的完整名称为“具有两个重复测量的两因素设计一元定量资料(主要疗效指标是血流速度)”。

4. 采用多次 t 检验进行多组间均值的两两比较

【例 6-19】 肥厚型心肌病患者心脏肌钙蛋白 T 基因变异及其相关临床表型分析。目的:探讨中国人群肥厚型心肌病(HCM)患者中心脏肌钙蛋白 T 基因变异的情况,并分析基因型与临床表型的关系。统计学分析:采用 SPSS 18.0 统计学软件,计数资料组间率的比较采用 χ^2 检验,计量资料以 $\bar{x} \pm s$ 表示,多组间均值比较采用单因素方差分析检验,组间两两比较采用 t 检验,以 $P < 0.05$ 为差异有统计学意义。

【实例解析】 原文中“计数资料”的正确表述应该为“定性资料”。原文作者在进行肥厚型心肌病患者心脏肌钙蛋白 T 基因变异及其相关临床表型分析时,并在统计学处理中谈及“计量资料以 $\bar{x} \pm s$ 表示,多组间均值比较采用单因素方差分析检验,组间两两比较采用 t 检验”,其中对多组间均值的两两比较时,原作者只是笼统提到“组间两两比较采用 t 检验”,故推测原作者很可能采用多次两样本 t 检验进行组间的两两比较。其实这种做法是不正确的,这样做会大大增加犯假阳性错误的概率。正确的做法是先对多组间均值进行单因素方差分析,如果差异有统计学意义,则可以选用 SNK- q 检验或 LSD- t 检验等进行两组间均值的比较。

5. 误用配对设计定量资料 t 检验处理具有一个重复测量的单因素设计定量资料

【例 6-20】 某研究者评估大剂量厄贝沙坦治疗慢性肾脏病(CKD)患者轻、中度蛋白尿的疗效与安全性。对 70 例患者采取大剂量(300~600mg/d)治疗 48 周。分别于治疗前、治疗后 12 周、治疗 24 周和治疗 48 周检测 Alb、TC、LDL-C 等生化指标。统计学处理:采用 SPSS 13.0 统计软件处理。计量资料以 $\bar{x} \pm s$ 表示,采用配对 t 检验。 $P < 0.05$ 表示差异有统计学意义。结论:大剂量组患者治疗前后各生化指标的变化中,Alb、TC、LDL-C 治疗前后均无明显变化(P 值均 > 0.05 ,见表 6-23)。

表 6-23 大剂量厄贝沙坦治疗慢性肾脏病患者前后各时点生化指标的变化($\bar{x} \pm s$)

组 别	完成例数(例)	Alb(g/L)	TC(mmol/L)	LDL-C(mmol/L)
治疗前	28	37.10 \pm 1.42	4.84 \pm 0.30	2.76 \pm 0.22
治疗后 12 周	22	38.06 \pm 1.19	4.97 \pm 0.51	2.77 \pm 0.27
治疗后 24 周	14	38.60 \pm 1.27	5.11 \pm 0.60	2.78 \pm 0.30
治疗后 48 周	6	40.61 \pm 2.44	5.02 \pm 0.45	2.84 \pm 0.26
P 值	0.496	0.980	0.995	

【实例解析】 研究者在用药前和用药后 3 个时间点上重复观测了 70 例慢性肾脏病患者的 Alb、TC、LDL-C。这样从每名患者身上就获得了某个指标的 5 个具有相关性的数据，因此“时间”是一个重复测量因素。此外，试验未涉及其他因素，故本资料为具有一个重复测量的单因素设计定量资料。研究者将某两个时间点上的数据放在一起，采用配对 *t* 检验进行分析，这种做法是错误的。这样做是随意割裂原本统一的整体设计。正确的做法是，采用具有一个重复测量的单因素设计定量资料方差分析来处理。

6. 误用配对设计定量资料 *t* 检验处理具有一个重复测量的两因素设计定量资料

【例 6-21】 研究者观察两种健康教育模式对等待尸体肾移植患者心理状况和营养状况的影响，探讨适合患者的健康教育模式。选择从 2009 年 9 月 1 日至 2010 年 8 月 30 日在某医院肾脏病中心确诊为慢性肾衰竭等待尸体肾移植的 125 例患者为研究对象，并将其随机分为对照组(62 例)和观察组(63 例)。对照组采用传统的健康教育，在住院期间进行常规术前教育。对观察组除给予常规术前教育外，还配备专职护士进行营养评估，建立个人档案，给予个性化饮食指导和移植相关知识教育，并跟踪随访；采用多样化教育形式，为患者提供与手术主管医生见面沟通的机会，与肾移植术后患者进行交流等。采用焦虑自评量表(SAS)、抑郁自评量表(SDS)对两组患者教育前后进行心理测评，见表 6-24。统计学处理：计量资料用 $\bar{x} \pm s$ 表示，采用配对 *t* 检验；计数资料采用 χ^2 检验。数据采用 SPSS 11.0 系统软件包处理。*P* < 0.05 表示差异有统计学意义。结论：健康教育后两组患者 SAS 和 SDS 评分比教育前均有下降，差异有统计学意义(*P* < 0.01)；相比对照组，观察组的 SAS 和 SDS 积分下降得更加明显(*P* < 0.01)。

表 6-24 接受两种健康教育模式的患者心理评分比较 ($\bar{x} \pm s$)

组 别	例数(男/女)	SAS(分)	SDS(分)
对照组	38/24		
教育前		57.89 ± 14.1	57.51 ± 9.39
教育后		47.05 ± 10.32 ^a	50.83 ± 10.12 ^a
观察组	41/22		
教育前		58.13 ± 12.02	57.18 ± 9.87
教育后		40.02 ± 9.05 ^{ad}	42.70 ± 10.01 ^{ad}

注：SAS: Zung 焦虑自评量表；SDS: Zung 抑郁自评量表；与同组教育前比较，^a代表 *P* < 0.01；与对照组教育后比较，^d代表 *P* < 0.01。

【实例解析】 本例作者将确诊为慢性肾衰竭等待尸体肾移植的 125 例患者随机分为对照组(62 例)和观察组(63 例)。对照组采用传统的健康教育，在住院期间进行常规术前教育；对观察组除给予常规术前教育外，还配备专职护士进行营养评估，建立个人档案，给予个性化饮食指导和移植相关知识教育，并跟踪随访；采用多样化教育形式。因此，教育方式是一个试验因素。另外，研究者在教育前后对研究者的心理评分及其他临床资料进行了测定，故观察时间是一个重复测量因素。因此，该资料应为具有一个重复测量的两因素设计定量资料。研究者采用配对 *t* 检验来分析两组患者在教育前后各指标取值的差异是否有统计学意义，并比较了两组患者 SAS 和 SDS 积分下降值间是否有差异(因原作者在此处未提及所采用的具体方法，只是在统计学处理方法处提及计量资料采用配对 *t* 检验方法，故推测作者可能仍然采用的是采用配对 *t* 检验)。这样做割裂了资料的整体设计，增大了结论犯假阳性错误的概率。合适的做法是，采用具有一个重复测量的两因素设计定量资料方差分析来处理，也可以以患者受教育前各指标的观察值为基础值，采用成组设计定量资料的一元协方差分析。

7. 误用配对秩和检验处理具有一个重复测量的两因素设计定量资料

【例 6-22】 某研究者欲评价在常规治疗基础上应用托伐普坦治疗心力衰竭(心衰)合并低钠血症的疗效和安全性。将多中心入选的 65 例心衰合并低钠血症患者随机、双盲分为托伐普坦组和安慰剂组。根据血钠情况,分别每日服用 15~60mg 托伐普坦片或安慰剂。观察患者第 1~7 天的体重变化情况,其中两组患者的体重减轻情况见表 6-25。统计分析:治疗后体重变化应采用配对秩和检验。结论:托伐普坦组体重下降优于安慰剂组($P<0.05$)。

表 6-25 两组治疗第 1~7 天体重减轻相对基线变化值

组 别	与基线比较的差值						
	第 1 天	第 2 天	第 3 天	第 4 天	第 5 天	第 6 天	第 7 天
安慰剂组(30 例)	—	0.30	0.38	0.51	0.47	0.31	0.47
托伐普坦组(35 例)	—	0.25	0.64 ^{ab}	0.82 ^{ab}	0.92 ^{ab}	1.01 ^{ab}	1.12 ^{ab}

注:与基线比较,^a 代表 $P<0.05$;与安慰剂组比较,^b 代表 $P<0.05$ 。

【实例解析】 本资料研究者将多中心入选的 65 例心衰合并低钠血症患者随机、双盲分为托伐普坦组和安慰剂组。两组受试者在试验条件上的差异在于所用药物不同,故“药物”是一个试验因素。研究者观察并比较服药前和服药后第 1~7 天的体重变化情况,所以观测时间也是一个因素,且为重复测量因素。因此,该资料为具有一个重复测量的两因素设计定量资料。研究者在文中笼统地提及“治疗后体重变化应选用配对秩和检验”。这种分析方法割裂了资料的完整性,这种做法不合适。合适的做法为可采用具有一个重复测量的两因素设计定量资料的方差分析进行处理。也可以患者服药前的体重为“基础值”,采用成组设计定量资料的协方差分析进行处理。从原文可知,原作者采用配对秩和检验比较两组患者治疗第 1~7 天体重减轻相对于基线的变化值是否有差异,故采用后一种方法较为合适。另外,作者在比较托伐普坦组与基线组每天的体重减轻情况时,采用的是配对秩和检验。由此可推测本资料中的体重变化情况不满足参数检验条件。基于此,可寻找合适的变量转换方法使转换后资料满足参数检验前提条件,然后对转换后的资料采用相应设计类型定量资料的方差分析进行处理。

8. 误用单因素 Kruskal-Wallis 秩和检验处理对照不全的两因素非平衡组合试验结果

【例 6-23】 某研究者欲探讨不同质子泵抑制剂(PPI)对大鼠非甾体类抗炎药(NSAID)所致小肠损伤是否具有保护作用及可能的保护机制。将 72 只 SD 大鼠随机分为空白对照组、模型对照组和奥美拉唑治疗组、埃索美拉唑治疗组、雷贝拉唑治疗组、兰索拉唑治疗组。除空白对照组外,其余各组予双氯芬酸 7.5mg/kg, d 灌胃,1 次/d,制备 NSAID 相关性小肠损伤大鼠模型。各治疗组分别予奥美拉唑 30mg/kg, d、埃索美拉唑 30mg/kg, d、兰索拉唑 45mg/kg, d、雷贝拉唑 15mg/kg, d,1 次/d 灌胃。连续给药 5d,处死后取小肠组织,观察其大体和病理学损伤变化。结果见表 6-26。统计学方法:应用 SPSS 11.5 软件进行统计分析,大体形态及病理组织学评分统计结果先用 Kruskal-Wallis 秩和检验并以均数(4 分位距)表示 EM(QR);组间两两比较用 Mann-Whitney 检验。

表 6-26 各组大鼠小肠组织大体及病理损伤积分[$n=12, M(Q_R)$]

评 分	组 别					
	空白对照组	模型对照	奥美拉唑	埃索美拉唑	雷贝拉唑	兰索拉唑
大体损伤积分	0.0(0.0)	4.00(4.00) ^a	2.75(1.75) ^b	1.62(1.00) ^{bc}	1.57(2.00) ^{bc}	1.77(1.50) ^{bc}
病理损伤积分	0.70(2.00)	2.35(1.75) ^a	1.60(2.50)	1.25(1.00) ^c	1.05(2.00) ^d	1.15(2.00) ^c

注:与空白对照组比较,^a 代表 $P<0.01$,^b 代表 $P<0.05$;与模型对照组比较,^c 代表 $P<0.05$,^d 代表 $P<0.01$ 。

【实例解析】 原统计表编制得不够规范，存在主谓颠倒现象。该资料的观测结果是大体损伤积分和病理损伤积分，为定量资料。其实验设计类型研究者将其视为单因素 6 水平设计，故采用了单因素 6 水平设计定量资料的秩和检验对其进行分析。该资料的“组别”并非只有一个实验因素，其中空白对照组和模型组相比较，其试验条件的差别在于是否建模，因此，是否建模是一个试验因素。另外，模型对照组、奥美拉唑组、埃索美拉唑组、雷贝拉唑组和兰索拉唑组之间的差别在于所用药物种类不同，模型组可以看成使用了“安慰剂”，所以药物种类是另一个试验因素。由此可知，该资料有两个因素，即是否建模和药物种类，其水平数分别为 2 和 5，故共有 10 个组合的实验条件，但实际上研究者只在其中的 6 种条件下进行了试验，所以本资料应为两因素非平衡组合试验定量资料。要对其进行正确分析，应该对组别进行合理拆分。注意：根据专业知识，通常应在模型基础上用药。拆分结果如下。

- 组合 1：空白对照组和模型对照组。
- 组合 2：模型对照组、奥美拉唑组、埃索美拉唑组、雷贝拉唑组、兰索拉唑组。

其中，组合 1 称为成组设计，其试验因素的名称是“建模与否”，即考察“建模与否”对观测指标的影响情况，对其分析应采用成组秩和检验（因研究者在文中指出结果变量不符合参数检验条件）；组合 2 称为单因素 5 水平设计，5 组动物都需要先建模，然后用不同的药物处理，第一组虽然没有用药，但可视为用了没有任何一种疗效的药物（如安慰剂），故这个因素称为“药物种类”，对其分析采用单因素 $k(k=5)$ 水平设计定量资料 Kruskal-Wallis 秩和检验。

9. 误用单因素方差分析处理对照不全的多因素非平衡组合试验定量资料

【例 6-24】 某研究者欲研究双歧杆菌分泌型粘附素对大鼠缺血再灌注(I/R)后肠黏膜屏障的防护作用。将雄性 72 只 SD 大鼠随机分为假手术组(24 只)、I/R 模型组(24 只)和粘附素预处理组(预处理组, 24 只)。建模成功后 6h 及 1d、4d、7d, 各组分别随机取 6 只大鼠剖杀, 观察小肠组织病理改变, 并检测各时间点血中二胺氧化酶(DAO)的活性和含量。见表 6-27。统计学处理: 应用 SPSS 13.0 统计学软件进行分析。计量数据采用 $\bar{x} \pm s$ 表示, 多组间计量资料比较采用单因素方差分析, 组间差异有统计学意义时采用 LSD- t 法进行两两比较。以 $P < 0.05$ 为差异有统计学意义。

表 6-27 各组大鼠造模后各时点的血浆二胺氧化酶水平 (U/mL, $\bar{x} \pm s$)

组 别	只 ^a	6h	1d	4d	7d
假手术组	6	1.18 ± 0.10	1.73 ± 0.15	1.91 ± 0.17	1.88 ± 0.15
I/R 模型组	6	1.70 ± 0.02 ^b	2.29 ± 0.05 ^b	2.41 ± 0.04 ^b	1.91 ± 0.05 ^b
预处理组	6	1.61 ± 0.04 ^{bc}	2.02 ± 0.05 ^{bc}	2.15 ± 0.09 ^{bc}	1.69 ± 0.03 ^{bc}

注: I/R 代表缺血再灌注; ^a代表每个时点的只数为 6 只; 与假手术组比较, ^b代表 $P < 0.05$; 与 I/R 模型组比较, ^c代表 $P < 0.05$ 。

【实例解析】 研究者将雄性 72 只 SD 大鼠随机分为三组: 假手术组、I/R 模型组和粘附素预处理组。假手术组和 I/R 模型组在试验条件上的差别在于是否造模, I/R 模型组和粘附素预处理组在试验条件上的差别在于是否使用粘附素进行预处理。另外, 对于本例建模成功后在各时点随机取 6 只大鼠剖杀, 故时间也是一个因素, 但并非为重复测量因素。因而该资料涉及三个实验因素, 即是否造模、是否使用粘附素预处理和时间因素, 分别包含 2 个、2 个和 4 个水平, 故其对应的实验设计类型为三因素某种设计类型。三因素多水平全面组合可形成 16 种实验条件, 而研究者仅在其中的 12 种条件下进行了试验, 所以该资料应为多因素非平衡组合试验一元定量资料。研究者将其视为单因素 3 水平设计一元定量资料, 采用单因素方差分析对资料进行处理。这种做法是不正确的! 肝炎发病率处理此资料的合适做法是依据专业知识和统计知识对资料进行合理拆分, 拆分结果如下: 组合 1 为假手术组、I/R 模型组; 组合 2 为 I/R 模型组、粘附素预处理组。

组合 1 和组合 2 均只涉及一个 2 水平的实验因素, 均为成组设计一元定量资料。在资料满足参

数检验的前提条件下,可选用成组设计一元定量资料 t 检验;若组间不满足方差齐性,但满足正态性可采用 t' 检验或秩和检验;若资料不满足正态性,可采用 Wilcoxon 秩和检验。

10. 误用单因素方差分析处理具有一个重复测量的多因素非平衡组合试验定量资料

【例 6-25】 原文题目:生长抑素和生长激素联用对兔重症急性胰腺炎肠黏膜屏障损伤的保护作用。目的:观察生长抑素(SS)和生长激素(GH)联合应用对重症急性胰腺炎(SAP)兔肠黏膜屏障损伤的保护作用,探讨其治疗 SAP 的意义。方法:将 36 只新西兰大白兔平均分为 3 组:SAP 模型组(SAP 组)、SS 治疗组(SS 组)以及 SS 和 GH 联合治疗组(SS + GH 组)。观察各组动物造模后第 6、12、24、48 小时血浆二胺氧化酶水平的变化。采用 SPSS 16.0 统计软件进行分析,组间比较采用单因素方差分析。结果见表 6-28。

表 6-28 各组重症急性胰腺炎大白兔血浆二胺氧化酶的变化情况(U/L, $\bar{x} \pm s$)

组 别	动 物 数	6h	12h	24h	48h
SAP 模型组	12	9.45 ± 4.83	18.15 ± 7.68	14.00 ± 3.52	7.67 ± 0.74
SS 组	12	5.87 ± 1.43	10.06 ± 1.32 ^a	8.74 ± 1.77 ^a	5.02 ± 0.95 ^a
SS + GH 组	12	5.70 ± 1.37	8.90 ± 0.98 ^a	4.61 ± 0.45 ^{ab}	3.75 ± 0.47 ^{ab}

注: SAP 为重症急性胰腺炎; SS 为生长抑素; GH 为生长素; ^a代表与 SAP 组比较, $P < 0.05$; ^b代表与 SAP 组比较, $P < 0.05$ 。

【实例解析】 研究者将 72 只新西兰大白兔平均分为 3 组,即 SAP 模型组(SAP 组)、SS 治疗组(SS 组)以及 SS 和 GH 联合治疗组(SS + GH 组)。对三组均进行了建模,其中 SS 组和 SS + GH 组相对于 SAP 组,又给予了额外的处理。SS 组在模型基础上由给予生长抑素(SS)治疗,所以是否给予生长抑素(SS)是一个试验因素。SS 和 GH 联合治疗组在给予生长抑素(SS)的基础上同时给予生长激素(GH),因此是否给予生长激素(GH)也是一个试验因素。另外,研究者在造模后第 6、12、24、48 小时重复观察每只动物血浆二胺氧化酶水平的变化,所以观察时间是一个具有重复测量的试验因素。由此可知,该资料共有三个实验因素,即是否给予生长抑素(SS)、是否给予生长激素(GH)和观察时间,其水平数分别为 2、2 和 4。这三个因素水平的全面组合应有 16 种实验条件,但实际上研究者只在其中的 12 种条件下进行了试验,所以本资料应为具有一个重复测量的多因素非平衡组合试验定量资料。研究者将其视为单因素 3 水平定量资料,采用单因素方差分析进行处理。这样做是不正确的,因为这样做无法准确反映组间的特点,且某些组间的差别可能是由多个因素造成的。正确的做法应为依据专业知识对其进行合理拆分,拆分结果如下:

组合 1: SAP 模型组和 SS 组。此组合仅涉及一个 2 水平的实验因素,即是否给予生长抑素(SS),故资料应为成组设计定量资料。

组合 2: SS 组和 SS + GH 组。该组合也只涉及一个两水平的实验因素,即是否给予给予生长激素(GH),应为成组设计定量资料。

以上两种组合均为成组设计定量资料,在资料满足参数检验的前提条件下,可选用成组设计定量资料 t 检验;若组间不满足方差齐性,但满足正态性时可采用 t' 检验或秩和检验;若资料不满足正态性时,可采用 Wilcoxon 秩和检验。

6.3.2 定性资料统计分析方法的合理选用

许多研究者在对定性资料进行统计分析时,不管资料取自何种设计,也不管变量类型如何,统统采用 χ^2 检验进行处理。这样做是不合理的。定性资料统计分析方法的合理选择要根据研究目的、试验设计方案和变量类型等因素进行。下面通过辨析内科学杂志中的一些样例,简单介绍定性资料统计分析方法的合理选用。

1. 一般 2 × 2 表资料的统计分析

所谓的一般 2 × 2 表资料是指由 2 行 2 列构成的资料。具体讲是指原因变量为 2 分类变量，结果变量也是 2 分类变量。对于这种资料可采用一般 χ^2 检验、校正 χ^2 检验和 Fisher 精确检验方法进行分析。

【例 6-26】 某研究者想了解广西城镇和农村居民对艾滋病主要传播途径的知晓情况。采用多层随机抽样方法对广西地区 18 ~ 74 岁的 866 名常住居民进行随机调查，其中 436 例城镇居民中有 63 例知道艾滋病的传播途径，而农村的 430 例 CKD 患者中仅 27 例知道，见表 6-29。问：该地区城镇与农村居民艾滋病转播途径知晓率差异有无统计学意义？

表 6-29 城镇和农村居民艾滋病传播途径知晓情况

地 区	例 数		
	知 道	不 知 道	合 计
城镇	63	373	436
农村	27	403	430
合计	90	776	866

【实例解析】 本例资料有 2 行 2 列，原因变量为地区，有 2 个水平；结果变量是艾滋病传播途径知晓情况。故属 2 × 2 一般四格表资料，可采用一般 χ^2 检验对比分析城镇和农村居民艾滋病传播途径的知晓率情况。SAS 处理程序如下，程序见 nkkysj6_26。

```
Data nkkysj6_26;
  Input r c f @@;
  Datalines;
1 1 63  1 2 373  2 1 27  2 2 403
;
run;

Proc freq;
Weight f;
Tables r*c /chisq expected;
Run;
```

SAS 程序说明：变量 r 表示地区，1 表示城镇，2 表示农村；变量 c 表示艾滋病传播途径知晓情况，1 表示知道，2 表示不知道；变量 f 表示发生的频数。在 FREQ 过程中，用 weight 语句定义 f 变量为列联表中的实际频数，用 tables 语句表示将建立以变量 r 为行变量，以变量 c 为列变量的二维列联表，即四格表。同时选用 expected 表示输出每个格中的理论频数，用 chisq 表示对四格表做卡方检验。

SAS 运行结果：

“r * c”表的统计量			
统计量	自由度	χ^2 值	概率
卡方	1	15.5190	< .0001
似然比卡方	1	15.9294	< .0001
连续校正卡方	1	14.6540	0.0001
Mantel-Haenszel 卡方	1	15.5010	< .0001
Phi 系数		0.1339	
列联系数		0.1327	
Cramer V 统计量		0.1339	

SAS 结果解释：本部分结果为 χ^2 检验的结果及 3 个分析行列变量的关联性统计量。首先是 χ^2 检验的结果，其中第一列列出了各种进行 χ^2 检验的方法，从上到下分别为卡方(基本公式计算法)、似然比法、连续性校正法、Mantel-Haenszel 法；第二列为自由度；第三列为各种方法计算所得的 χ^2 值；第四列给出了 χ^2 值所对应的概率值(P 值)。根据资料所具备的不同条件可选择不同的结果。本例总例数大于 40，所有的理论频数均大于 5(最小的理论频数为 44.688)，所以选用基本公式法。即 $\chi^2 = 15.5190$, $P < 0.0001$ ，说明城镇和农村地区艾滋病传播途径知晓情况有差别，城镇居民较农村居民的知晓率较高。

χ^2 检验结果的下方列出的 3 个分析行列变量的关联性统计量，分别为 Phi 系数、列联系数和 Cramer V 统计量，它们的值均在 $-1 \sim +1$ 之间，绝对值越大，说明行列变量的关联性越密切。

2. 配对研究设计 2 × 2 表资料统计分析

所谓配对研究设计 2 × 2 表资料是受试者指按照配对原则分别接受两种不同的处理方法，每种处理方法的结果都可分为“阳性”和“阴性”两种，清点出两种处理方法同时判为阳性、阴性的频数以及它们结果不一致的频数，将结果列成配对设计的 2 × 2 表形式，其资料表达形式如表 6-30 所示。

表 6-30 配对研究设计 2 × 2 表

属 性 A	频 数		
	属性 B: +	属性 B: -	合 计
+	<i>a</i>	<i>b</i>	<i>a + b</i>
-	<i>c</i>	<i>d</i>	<i>c + d</i>
合计	<i>a + c</i>	<i>b + d</i>	<i>N</i>

【例 6-27】 某研究者在分析口腔和胃幽门螺杆菌(Hp)的感染情况时，对 129 例胃幽门螺杆菌(Hp)感染者同步采用¹³C/¹⁴C 尿素呼气试验法和螺旋杆菌抗原技术法进行检测，检测结果见表 6-31，试比较两种方法的检测结果是否一致。

表 6-31 两种方法检测结果比较(例)

螺旋杆菌抗原技术法	例 数		
	* : 阳性	阴 性	合 计
阳性	39	59	98
阴性	5	26	31
合计	44	85	129

注：* 代表“¹³C/¹⁴C 尿素呼气试验法”。

【实例解析】 本例研究者采用螺旋杆菌抗原技术法与¹³C/¹⁴C 尿素呼气试验法同步对 129 例胃幽门螺杆菌(Hp)感染患者进行检测，故属于配对设计四格表资料。对于特设金标准的配对设计四格表资料，其分析目的不同，所用的检验方法不同。若分析目的是为了比较两种检测方法不一致部分的频数之间的差别是否具有统计学意义，则可选用配对设计定性资料的 χ^2 检验，即 McNemar χ^2 检验；若研究目的是为了检验两种检验方法检测结果是否一致，则进行一致性检验，即采用 Kappa 检验进行分析。

现假定¹³C/¹⁴C 尿素呼气试验法是金标准，则本资料是特设金标准的配对设计四格表资料。并且原文的目的是为了检验¹³C/¹⁴C 尿素呼气试验法和螺旋杆菌抗原技术法检测结果是否一致，故选用 Kappa 检验方法处理资料。SAS 处理程序如下，程序见 nkysj6_27。

```
Data nkkysj6_27;  
do r=1 to 2;  
do c=1 to 2;  
Input f @@;  
output;  
end;  
end;  
Datalines;  
39 59 5 26  
;  
run;  
Proc freq;  
weight f;  
tables r*c /agree;  
Run;
```

SAS 程序说明：在 tables 语句中使用了 agree 选项，表示将在结果中输出 McNemar 检验和一致性检验的结果。

SAS 运行结果：

“r * c”表的统计量	
McNemar 检验	
统计量(S)	45.5625
自由度	1
Pr > S	< .0001
简单 Kappa 系数	
Kappa	0.1483
渐近标准误差	0.0561
95% 置信下限	0.0383
95% 置信上限	0.2584
样本大小 = 129	

SAS 结果解释：配对四格表资料 χ^2 检验的结果除了列联表的内容外，还会出现上述两部分的结果。一部分是 McNemar 检验的统计量、自由度和 P 值，本例 $\chi^2 = 45.5625$, $P < 0.0001$ ，说明螺旋杆菌抗原技术法和 $^{13}\text{C}/^{14}\text{C}$ 尿素呼气试验法对胃幽门螺杆菌 (Hp) 阳性检出数的差异有统计学意义，但本例的研究目的是考察两种检验方法的检验结果是否一致，故选择第二部分结果，即一致性检验结果。第二部分是一致性检验的 Kappa 值、渐近标准误和 95% 可信区间。本例 95% 可信区间为 (0.0383, 0.2584)，该区间不包含 0，说明 $^{13}\text{C}/^{14}\text{C}$ 尿素呼气试验法和螺旋杆菌抗原技术法对胃幽门螺杆菌 (Hp) 感染者检测结果具有一致性(注意：若 Kappa = 0，则意味着两种测定方法测定的结果不一致)。值得一提的是：在本例中，虽然统计结论是两种测定方法测定的结果具有一致性，但实际的一致率 $65/129 = 50.39\%$ 还是相当低的。此时，需要结合专业知识，做出专业结论。

3.2 × K 列联表资料统计分析

所谓 2 × K 列联表资料是指 2 行 K 列的列联表资料。具体说就是原因变量为 2 值变量，结果变量为多值变量的列联表。如果结果变量为多值名义变量，则可按双向无序的 $R \times C$ 表资料进行分析，采用一般 χ^2 检验或 Fisher 精确检验；如果结果变量为多值有序变量，则可按结果变量为单向有序的 $R \times C$ 表资料进行分析，采用秩和检验或 Ridit 分析等方法。

【例 6-28】 某研究者欲研究胸腔镜辅助微创与导管射频消融治疗持续性心房颤动(房颤)的临床疗效。对北京某医院 172 例持续性房颤患者进行射频消融,其中 83 例行胸腔镜辅助微创射频消融(微创组),另 89 例行环同侧肺静脉左心房线性消融(导管组)。对两组患者的术后心律情况进行统计,见表 6-32。

表 6-32 两组术后即刻心律种类情况(例)

组 别	例 数			
	窦 性	房 颤	房 扑	起 搏
微创组	38	36	5	4
导管组	29	36	24	0

表 6-32 中的原因变量(组别)是名义变量(2 值),结果变量(心律种类)是名义分类变量,因而可按双向无序的 $R \times C$ 表资料进行分析,采用一般 χ^2 检验或 Fisher 精确检验。本例中因有小于 1 的理论数,故可用 Fisher 精确检验法。SAS 运行程序如下,程序见 nkkysj6_28。

```
data nkkysj6_28;
do r=1 to 2;
do c=1 to 4;
input f @@;
output;
end;
end;
datalines;
38 36 5 4
29 36 24 0
;
run;
proc freq;
weight f;
tables r*c /fisher;
run;
```

SAS 运行结果:

“r * c”表的统计量

统计量	自由度	值	概率
卡方	3	17.4692	0.0006
似然比卡方	3	20.0888	0.0002
Mantel-Haenszel 卡方	1	4.1297	0.0421
Phi 系数		0.3187	
列联系数		0.3036	
Cramer V 统计量		0.3187	

WARNING: 25% 的单元格的期望计数比 5 小。卡方可能不是有效检验。

Fisher 精确检验

表概率(P)	1.265E - 06
Pr < = P	2.764E - 04

样本大小 = 172

SAS 结果解释：因本实例中有小于 1 的理论频数，故选择 Fisher 精确检验法的运行结果。本例精确检验结果的概率 $P = 0.0002764$ ，故可知微创组和导管组两组患者的术后心律情况的差别具有统计学意义。

4. $K \times 2$ 列联表资料统计分析

所谓 $K \times 2$ 列联表资料是指 K 行 2 列的列联表资料，具体讲是指原因变量有 K 个取值的变量，而结果为二值变量的列联表。如果原因变量为多值名义变量则可按双向无序的 $R \times C$ 表资料进行分析，采用一般 χ^2 检验或 Fisher 精确检验。如果原因变量为多值有序变量，则要根据分析目的选择合适的检验方法：若要分析原因变量间的频数分布是否不同，则可用一般 χ^2 检验或 Fisher 精确检验；若要分析第一列（或第二列）上的频率变化是否呈线性趋势，则要选用线性趋势检验方法。对于检验方法的具体计算过程，这里不做详述。

【例 6-29】 某研究者对腹膜炎致病菌的耐药性进行研究，其中革兰氏阴性杆菌的耐药情况见表 6-33。试对比分析各种抗生素的耐药情况。

表 6-33 革兰氏阴性杆菌的耐药情况(株数)

抗 生 素	例 数		
	耐药性：敏感	耐 药	合 计
亚胺培南	8	0	8
头孢他啶	7	6	13
环丙沙星	10	3	13
头孢吡肟	2	3	5
庆大霉素	13	3	16
复方 SMZ	8	3	11
氨苄西林	1	8	9
头孢哌酮舒巴坦	5	2	7
哌拉西林	8	5	13

【实例解析】 表 6-33 中的原因变量抗生素的种类，为多值名义分类变量，结果变量为耐药性，因此该资料视为 $K \times 2$ 列联表资料。本资料的研究目的为比较不同抗生素的耐药情况，故选用一般 χ^2 检验或 Fisher 精确检验分析方法。SAS 程序名为 nkkysj6_29。

```
data nkkysj6_29;
do r=1 to 9;
do c=1 to 2;
input f @@;
output;
end;
end;
datalines;
8 0 7 6 10 3 2 3 13 3 8 3 1 8 5 2 8 5
;
run;
proc freq;
weight f;
tables r*c /fisher;
run;
```

SAS 运行结果：

FREQ PROCEDURE
“r * c”表的统计量

统计量	自由度	值	概率
卡方	8	21.1054	0.0069
似然比卡方	8	23.6708	0.0026
Mantel-Haenszel 卡方	1	2.2809	0.1310
Phi 系数		0.4713	
列联系数		0.4264	
Cramer V 统计量		0.4713	

WARNING: 56% 的单元格的期望计数比 5 小。卡方可能不是有效检验。

Fisher 精确检验

表概率(P)	2.918E - 09
Pr < = P	0.0053

样本大小 = 95

SAS 结果解释：因本资料的研究目的为比较不同抗生素的耐药情况，且含有小于 1 的理论数，故选用 Fisher 精确检验方法。本例中 $P = 0.0053$ ，由此可知这 9 种抗生素的耐药情况总体上有差别，具体哪几种不同需采用卡方分割法进行检验，在此不再详述。

【例 6-30】 某研究者探讨核黄素缺乏与食管鳞癌发生的关系，采用 ELISA 法检测山西长治地区食管鳞癌患者(445 例)、长治当地健康人群(689 例)及健康林州移民(347 例)3 组人群的血浆中核黄素水平。为真实反映各组的检测结果，该研究者对 3 组性别构成情况分别进行了比较。结果见表 6-34。

表 6-34 3 组人群的性别构成情况比较[例(构成比, %)]

组 别	男 性	女 性	合 计
当地食管鳞癌患者	362	83	445
当地健康对照组	530	159	689
健康移民组	270	77	347
合计	1162	319	1481

【实例解析】 由表 6-34 可知，该资料的类型为 $K \times 2$ 列联表资料。本资料的研究目的为分析比较不同组别人群与性别之间是否有关联，故可选用一般 χ^2 检验对资料进行处理。SAS 运行程序如下，程序见 nkkysj6_30_1。

```
data nkkysj6_30_1;
do r=1 to 3;
do c=1 to 2;
input f @@;
output;
end;
end;
datalines;
362 83 530 159 270 77
;
```

```
run;
proc freq;
  weight f;
  tables r*c /cmh;
run;
```

SAS 程序说明：在处理行 × 列表资料时，tables 语句后面的选项可以不用 chisq，直接用 cmh。
SAS 运行结果：

“r * c”的汇总统计量				
Cochran-Mantel-Haenszel 统计量(基于表得分)				
统计量	对立假设	自由度	值	概率
1	非零相关	1	1.7184	0.1899
2	行均值得分差值	2	3.2443	0.1975
3	一般关联	2	3.2443	0.1975
总样本大小 = 1481				

SAS 结果解释：本例资料的类型为 $K \times 2$ 列联表资料。行和列均为无序资料，故欲分析比较不同组别人群的性别构成情况，可选用一般 χ^2 检验对其进行分析。分析结果中包含 3 个统计量，其中第 1 个统计量为非零相关，表示行变量和列变量为非零相关，可用于双向有序资料；第 2 个统计量为行均值得分差值，可用于列变量(结果变量) 为有序变量的资料；第 3 个统计量为一般关联，表示行变量和列变量为一般关联，可用于双向无序资料。因本例为双向无序资料，故选择第 3 个统计量，统计量大小为 $\chi^2_{CMH} = 3.2443$ ， $P = 0.1975 > 0.05$ ，故可认为 3 组不同人群与性别之间无关联，其专业结论是：3 组不同人群在男、女性别构成或比例上是接近相等的。

若本例的研究目的仅为了解 3 组人群的性别分布情况，可选用双向无序资料一般卡方检验(避免从 CMH 检验给出的三种检验结果中选择其中之一)，其 SAS 处理程序如下，程序见 nkkysj6_30_2。

```
data nkkysj6_30_2;
do r=1 to 3;
do c=1 to 2;
  input f @@;
  output;
end;
end;
datalines;
362 83 530 159 270 77
;
```

```
run;
proc freq;
  weight f;
  tables r*c /chisq;
run;
```

SAS 运行结果：

“r * c”表的统计量			
统计量	自由度	值	概率
卡方	2	3.2465	0.1973
似然比卡方	2	3.3068	0.1914

Mantel-Haenszel 卡方	1	1.7184	0.1899
Phi 系数		0.0468	
列联系数		0.0468	
Cramer V 统计量		0.0468	

样本大小 = 1481

SAS 结果解释：本例的检验检验结果为： $\chi^2 = 3.2465, P = 0.1973$ ，由此可知，3 组人群中性别的频数分布之间的差别无统计学意义，即男女性别构成基本相等。

5. 双向无序 $R \times C$ 列联表资料的统计分析

所谓的双向无序 $R \times C$ 列联表资料是指在二维列联表中，两个变量都是名义分类变量，并且这两个名义变量分别有 R 个和 C 个可能取值，由此排列成的 $R \times C$ 表称为双向无序 $R \times C$ 列联表。处理该类资料可用一般 χ^2 检验或 Fisher 精确检验。

【例 6-31】 某研究者回顾性收集 160 例心肌病患者的临床资料，分成 3 组进行分析。其中心尖肥厚型心肌病(AHCM)组 41 例，非梗阻性肥厚型心肌病(NOHCM)组 52 例，梗阻性肥厚型心肌病(OHCM)组 67 例。三组心肌病患者的临床症状见表 6-35。

表 6-35 三组心肌病患者的临床症状

组 别	例 数				
	劳力性呼吸困难	胸 痛	晕 厥	心 悸	合 计
AHCM 组	12	13	2	23	41
NOHCM 组	28	17	6	16	52
OHCM 组	46	20	21	27	67
合计	86	50	29	66	160

【实例解析】 从表 6-35 可知，本例资料为双向无序 $R \times C$ 列联表资料，研究目的为比较不同组别的心肌病患者的临床症状频数分布情况，故可用一般 χ^2 检验或 Fisher 精确检验对资料进行统计分析。SAS 运行程序如下，程序见 nkkysj6_31。

```
data nkkysj6_31;
do r=1 to 3;
do c=1 to 4;
input f @@;
output;
end;
end;
datalines;
12 13 2 23
28 17 6 16
46 20 21 27
86 50 29 66
;
run;
proc freq;
weight f;
tables r*c /chisq;
run;
```

SAS 运行结果：

“r * c”表的统计量

统计量	自由度	值	概率
卡方	6	9.2354	0.1608
似然比卡方	6	9.0205	0.1724
Mantel-Haenszel 卡方	1	1.4988	0.2209
Phi 系数		0.1414	
列联系数		0.1400	
Cramer V 统计量		0.1000	

样本大小 = 462

SAS 结果解释：因为本例为双向无序资料，所以采用一般卡方检验的结果， $\chi^2 = 0.1608$ ， $P = 0.1618 > 0.05$ ，故可认为三组心脏病患者的临床症状的频数分布相同。

6. 原因变量为有序变量、结果变量为二值变量的单向有序资料分析

对于原因变量为多值有序变量、结果变量为二值变量的资料，其统计分析方法选用应视研究目的而定。如果要分析原因变量间频数分布是否不同，则可选用一般 χ^2 检验或 Fisher 精确检验；如果要分析第一列或第二列上的频数变化是否呈线性趋势时，则选用线性趋势检验。

【例 6-32】 某研究者欲比较不同心功能 Killip 分级的急性心肌梗死患者的病死情况，对 2002 年 1 月—2010 年 2 月期间在某院住院的不同心功能 Killip 分级患者的急性心肌梗死病死率进行了分析。其病死情况见表 6-36。

表 6-36 不同心功能 Killip 分级急性心肌梗死患者的病死情况

心功能 Killip 分级	死 亡 组	存 活 组	合 计
I 级	16	140	156
II 级	18	166	184
III 级	30	45	75
IV 级	33	51	84
合计	97	402	499

【实例辨析】 表 6-36 中的原因变量为心功能 Killip 分级(等级资料)，结果变量为病死情况，其取值为死亡、存活，为名义变量。故该资料属于原因变量为有序变量的单向有序资料。本资料研究者想比较不同心功能 Killip 分级的急性心肌梗死患者的病死情况，故可将该资料视为双向无序的 $R \times C$ 列联表资料，选用一般的卡方检验进行分析。SAS 运行程序如下，程序见 nkkysj6_32。

```
data nkkysj6_32;
do r=1 to 4;
do c=1 to 2;
input f @@;
output;
end;
end;
datalines;
16 140 18 166 30 45 33 51
;
run;
proc freq;
weight f;
tables r*c /chisq;
run;
```

SAS 运行结果：

“r * c”表的统计量			
统计量	自由度	值	概率
卡方	3	60.7303	<.0001
似然比卡方	3	56.9908	<.0001
Mantel-Haenszel 卡方	1	45.9644	<.0001
Phi 系数		0.3489	
列联系数		0.3294	
Cramer V 统计量		0.3489	
样本大小 = 499			

SAS 结果解释：本例资料的卡方值为 60.7303, $P < 0.0001$ ，可知不同心功能 Killip 分级患者的急性心肌梗死病死情况不同，由原始数据可初步认为：其中Ⅲ级和Ⅳ级较Ⅰ级和Ⅱ级的病死率高。值得一提的是：若研究者希望回答“死亡率是否会随着心功能 Killip 分级上升而呈线性上升趋势”，可在上面的 SAS 程序中，将语句“tables r * c /chisq;”改为“tables r * c /trend;”即可。

7. 结果变量为有序变量的单向有序 $R \times C$ 列联表资料分析

结果变量为有序变量的单向有序 $R \times C$ 列联表资料是指表中仅结果变量的取值为有序的，而原因变量是无序的。其统计分析可根据研究目的选用秩和检验、Ridit 分析以及有序变量的 logistic 回归分析。

【例 6-33】 某研究者欲探讨不同药物对大鼠小肠组织转录因子红细胞系-2p45 相关因子-2 (Nrf2)表达水平的影响。将 50 只 SD 大鼠随机分为模型对照组、奥美拉唑治疗组、埃索美拉唑治疗组、雷贝拉唑治疗组和兰索拉唑治疗组。通过 Western 印迹检测、实时 PCR 对各组 Nrf2 的表达水平(分为四级,分别为 -, +, ++)进行了分析,结果见表 6-37。试比较分析各组 Nrf2 表达间的差别是否具有统计学意义。

【实例辨析】 表 6-37 中的原因变量为不同用药组(名义变量),结果变量为心功能 Nrf2 表达水平,其取值为“-、+、++”,为有序变量。因此,该资料可按结果变量为有序变量的单向有序 $R \times C$ 列联表资料进行分析,根据研究目的可选用一般秩和检验和 Ridit 分析,而不能选用一般的 χ^2 检验。因为一般的 χ^2 检验体现不出变量的有序性,所得到的结论只能反映各组的频数分布是否相同,而不能得出各组 Nrf2 表达水平间差别是否有统计学意义的结论。本资料研究者欲比较各组的 Nrf2 表达水平间是否有差别,故选用秩和检验。SAS 运行程序如下,程序见 nkkysj6_33。

表 6-37 各组免疫组化小肠组织转录因子红细胞系-2p45 相关因子-2 (Nrf2) 表达水平

组 别	例 数			合 计
	* : -	+	++	
模型对照组	9	1	0	10
奥美拉唑组	8	1	1	10
埃索美拉唑组	8	1	1	10
雷贝拉唑组	5	3	2	10
兰索拉唑组	6	2	2	10

注：* 代表“Nrf2 表达水平”。

```
data nkkysj6_33;
do a=1 to 5;
do b=1 to 3;
input f @@;
output;
end;
end;
datalines;
9 1 0
8 1 1
8 1 1
5 3 2
6 2 2
;
proc nparlway wilcoxon data=nkkysj6_33;
class a;
var b;
freq f;
run;
```

SAS 运行结果：

The NPAR1WAY Procedure					
Wilcoxon Scores(Rank Sums)for Variable b					
Classified by Variable a					
a	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	10	207.0	255.0	32.496154	20.70
2	10	236.0	255.0	32.496154	23.60
3	10	236.0	255.0	32.496154	23.60
4	10	309.0	255.0	32.496154	30.90
5	10	287.0	255.0	32.496154	28.70
Average scores were used for ties.					
Kruskal-Wallis Test					
Chi-Square		5.2773			
DF		4			
Pr > Chi-Square		0.2600			

SAS 结果解释：其中第一部分结果为各组秩和的打分情况。其中，各组的平均秩和为：模型对照组为 20.70，奥美拉唑组为 23.60，埃索美拉唑组为 23.60，雷贝拉唑组为 30.90，兰索拉唑组为 28.70。第二部分结果是对 5 组的表达分级进行的 Kruskal-Wallis 检验，因 $H_c \approx \chi^2 = 5.2773$ ， $P = 0.2600$ ，故 5 个组的表达分级之间的差别无统计学意义，即 5 组表达近似相同。

【例 6-34】 某研究者在对 521 例大肠癌患者的临床特征进行分析时，对不同性别大肠癌患者的年龄段分布情况进行了统计。见表 6-38，试比较不同性别大肠癌患者的年龄段分布是否相同。

表 6-38 不同性别大肠癌患者的年龄段分布情况[例(构成比，%)]

性 别	例 数	≤30 岁	31~40 岁	41~50 岁	51~60 岁	61~70 岁	>70 岁
男性	295	7(2.4)	26(8.8)	51(17.3)	64(21.7)	60(20.3)	87(29.5)
女性	226	3(1.3)	19(8.4)	35(15.5)	68(30.1)	57(25.2)	44(19.5)
合计	521	10	45	86	132	117	131

【实例辨析】 在表 6-38 中, 原因变量(性别)是名义变量(2 值), 结果变量(年龄段)是有序变量, 因而该资料属于结果变量为有序变量的单向有序列联表资料。对于此类资料, 研究目的不同, 选择的统计检验方法不同。就本例来说, 当研究目的为研究两组大肠癌患者的年龄段分布是否相同时, 可采用一般 χ^2 检验, 此时可将其视为双向无序列联表资料; 但当研究目的为比较男性和女性两组大肠癌患者平均年龄之间的差别是否有统计学意义时, 则需选用秩和检验。因此, 本例资料根据其研究目的, 应选用一般 χ^2 检验来分析比较不同性别大肠癌患者的年龄段分布情况。SAS 程序名为 nkkysj6_34。

```
data nkkysj6_34;
do r=1 to 2;
do c=1 to 6;
input f @@;
output;
end;
end;
datalines;
7 26 51 64 60 87 3 19 35
68 57 44
;
run;
proc freq;
weight f;
tables r*c /chisq;
run;
```

SAS 运行结果:

“r * c”表的统计量			
统计量	自由度	值	概率
卡方	5	11.0336	0.0507
似然比卡方	5	11.1458	0.0486
Mantel-Haenszel 卡方	1	0.6274	0.4283
Phi 系数		0.1455	
列联系数		0.1440	
Cramer V 统计量		0.1455	
样本大小 = 521			

SAS 结果解释: 本例中的 $\chi^2 = 11.0336$, $P = 0.0507$, 故可知不同性别大肠癌患者的年龄段分布情况相同。

【例 6-35】 某研究者在原部分贲门肌切开的基础上行改良外置幽门部分结扎术, 并与传统幽门半缝扎术及内置幽门部分缝扎术进行比较。将 80 只体重相近的健康雄性 Wistar 大鼠随机均分为甲、乙、丙、丁组。甲组行传统贲门肌部分切开联合半幽门缝扎术, 乙组行贲门肌部分切开联合内置幽门部分缝扎术, 丙组行贲门肌部分切开联合外置幽门部分结扎术, 丁组为空白手术对照组。各组食管大体标本黏膜表现分级与病理分级情况一致, 见表 6-39。比较分析各组食管大体标本黏膜表现分级间的差异。

表 6-39 各组食管大体标本和病理检查黏膜表现分级情况(例)

组别	鼠数	正 常	轻 度	中 度	重 度
甲组	20	6	4	6	4
乙组	20	2	6	10	2
丙组	20	2	5	10	3
丁组	20	20	0	0	0

【实例辨析】 本例资料中的原因变量(组别)是名义分类变量,结果变量(病理分级)是有序变量,因而属于结果变量为有序变量的单向有序列联表资料。因一般 χ^2 检验与变量的有序性没有联系,用一般 χ^2 检验进行分析,得到的结论是 4 组病理分级的频数分布是否相同,而不是 4 组病理分级的差别是否有统计学意义。要比较 4 组大鼠病理分级间有无差别,可选用秩和检验。SAS 运行程序如下,程序见 nkkysj6_35。

```
data nkkysj6_35;
do a=1 to 4;
do b=1 to 4;
input f @@;
output;
end;
end;
datalines;
6 4 6 4
2 6 10 2
2 5 10 3
20 0 0 0
;
run;
proc npar1way wilcoxon data=nkkysj6_35;
class a;
var b;
freq f;
run;
```

SAS 运行结果:

a	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	20	900.0	810.0	85.621333	45.00
2	20	996.0	810.0	85.621333	49.80
3	20	1034.0	810.0	85.621333	51.70
4	20	310.0	810.0	85.621333	15.50

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square	35.0776
DF	3
Pr > Chi-Square	<.0001

SAS 结果解释: 第一部分结果为各组秩和的打分情况,其中各组的平均秩和是,甲组为 45.00,乙组为 49.80,丙组为 51.70,丁组为 15.50。第二部分结果是对 4 个组的黏膜表现分级进行的 Kruskal-Wallis 检验,因 $H_c \approx \chi^2 = 35.0776, P < 0.0001$,故 4 个组的黏膜表现分级之间的差别具有统计学意义,即 4 种方法黏膜表现分级与病理分级情况是不一致的。

8. 双向有序属性不同的 $R \times C$ 表资料及其统计分析

当 $R \times C$ 表中的两个定性变量,即原因变量和结果变量都是有序变量,并且它们的性质不同,这样的列联表资料称为双向有序属性不同的 $R \times C$ 表资料。

对于双向有序且属性不同的 $R \times C$ 列联表资料所采用的统计方法不能一概而论，应视具体的分析目的而定。一般来说，有 4 个可能的分析目的：

(1) 当分析目的只关心各组结果变量取值之间的差别是否具有统计学意义时，原因变量的有序性就变得无关紧要了，可将此时的“双向有序 $R \times C$ 列联表资料”视为结果变量为有序变量的单向有序 $R \times C$ 列联表资料，可以选择的统计方法有秩和检验，Ridit 分析和有序变量的 logistic 回归分析。

(2) 当分析目的为考察原因变量与结果变量之间是否存在相关关系时，可用 Spearman 秩相关分析。

(3) 当两个有序变量之间存在的相关关系有统计学意义，研究者希望进一步了解这两个有序变量之间的变化关系是直线关系还是呈某种曲线关系，此时宜选用线性趋势检验。

(4) 当分析目的只是希望考察各行上的频数分布是否相同，此时将其视为双向无序的 $R \times C$ 列联表资料，可根据资料具备的前提条件，选用一般 χ^2 检验或 Fisher 精确检验。

【例 6-36】 某研究者欲研究不同溃疡性结肠炎疾病活动度(轻度、中度和重度)间 γ -干扰素的表达强度。应用免疫组织化学染色法(SABC 法)检测并比较各组结肠黏膜固有层内单个核细胞中的 γ -干扰素的表达水平。结论：三组间结肠黏膜固有层内单个核细胞中 γ -干扰素的表达水平有统计学差异($\chi^2 = 11.628$, $P = 0.020$)，其表达强度随疾病活动度加重而升高，见表 6-40。

表 6-40 不同溃疡性结肠炎疾病活动度间 γ -干扰素表达强度的比较(例)

分 组	-	+	++	+++
轻度活动	0	7	2	0
中度活动	0	3	3	8
重度活动	0	3	3	1

【实例辨析】 表 6-40 中的原因变量为不同溃疡性结肠炎疾病的活动度(等级资料)，结果变量为 γ -干扰素表达强度，其取值为 -、+、++ 和 +++，为等级资料。故该资料为双向有序且属性不同的 $R \times C$ 列联表资料。

由此可知，本实例的研究目的是比较不同溃疡性结肠炎疾病活动度间 γ -干扰素的表达强度，因此宜采用 Kruskal-Wallis 秩和检验。如果用一般卡方检验对其进行分析，所得结论则为三种溃疡性结肠炎活动度 γ -干扰素表达强度的频数分布是否不同，而不能得出三种溃疡性结肠炎活动度 γ -干扰素表达强度间的差别是否有统计学意义的结论。SAS 运行程序如下，程序见 nkkysj6_36。

```
data nkkysj6_36;
do a=1 to 3;
do b=1 to 4;
input f @@;
output;
end;
end;
datalines;
0 7 2 0
0 3 3 8
0 3 3 1
;
run;
proc npar1way wilcoxon data =nkkysj6_36;
class a;
var b;
run;
```

SAS 运行结果：

a	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	9	84.00	139.50	20.653170	9.333333
2	14	281.50	217.00	22.484324	20.107143
3	7	99.50	108.50	19.062013	14.214286

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square	9.6147
DF	2
Pr > Chi-Square	0.0082

SAS 结果解释：第一部分结果为各种溃疡性结肠炎疾病活动度的秩和打分情况，其中各种活动度的平均秩和是，轻度为 9.333333，中度为 20.107143，重度为 14.214286。第二部分结果是对三种活动度的 γ -干扰素表达强度进行的 Kruskal-Wallis 检验，因 $H_c \approx \chi^2 = 9.6147, P = 0.0082$ ，故溃疡性结肠炎疾病活动度的 γ -干扰素表达强度之间有统计学意义，尚不能认为三种溃疡性结肠炎活动度 γ -干扰素表达强度无差异。

【例 6-37】 某研究者欲研究不同年龄组的肥胖情况，选取 2007—2009 年某省社区自然人群 2285 名为研究对象进行回顾性研究。结果见表 6-41。

表 6-41 不同年龄人群的体重分布情况[例(%)]

年 龄	正 常	过 轻	超 重	肥 胖
20 ~ 39 岁	280(45.5)	10(1.6)	233(37.8)	93(15.1)
40 ~ 59 岁	359(27.4)	9(0.7)	722(55.2)	218(16.7)
≥60 岁	113(31.3)	2(0.6)	184(51.0)	62(17.2)
合计	752(32.9)	21(0.9)	1139(49.8)	373(16.3)

【实例辨析】 由表 6-41 可知，原因变量为年龄组(等级资料)，结果变量为体重情况，其取值为正常、过轻、超重和肥胖，为等级资料。故该资料为双向有序且属性不同的 $R \times C$ 列联表资料。但分析目的只是希望考察各年龄组的体重分布情况是否相同，此时将其视为双向无序的 $R \times C$ 列联表资料，可根据资料具备的前提条件，选用一般 χ^2 检验或 Fisher 精确检验。SAS 运行程序如下，程序见 nkkysj6_37。

```
data nkkysj6_37;
do r=1 to 3;
do c=1 to 4;
input f @@;
output;
end;
end;
datalines;
280 10 233 93
359 9 722 218
113 2 184 62
752 21 1139 373
;
run;
proc freq;
weight f;
tables r*c /chisq;
run;
```

SAS 运行结果：

“r * c”表的统计量			
统计量	自由度	值	概率
卡方	6	32.1663	<.0001
似然比卡方	6	32.6588	<.0001
Mantel-Haenszel 卡方	1	14.5666	0.0001
Phi 系数		0.0839	
列联系数		0.0836	
Cramer V 统计量		0.0593	
样本大小 = 4570			

SAS 结果解释:本例中的 $\chi^2 = 32.1663$, $P < 0.0001$, 有统计学差异, 可知各年龄组的体重分布情况不同。

9. 双向有序属性相同的 $R \times C$ 列联表资料及其统计分析

当行变量与列变量的性质相同且取值的水平数及含义也相同时, 这样的资料称为双向有序属性相同的 $R \times C$ 列联表资料。双向有序且属性相同的 $R \times C$ 列联表资料是一个“方形”列联表, 实际上是配对设计 2×2 列联表资料的“扩大”。只不过对于配对设计 2×2 列联表资料, 人们更关心的是两种检测方法检测结果之间的差别是否具有统计学意义, 而双向有序且属性相同的 $R \times C$ 列联表资料的研究目的大多是检验行变量与列变量的检测结果间是否具有有一致性, 故常用的统计分析方法称为一致性检验或 Kappa 检验。

【例 6-38】 某研究者对 150 名冠心病患者用两种方法检查室壁收缩运动的情况, 检测结果见表 6-42。试分析两种方法测定结果有无一致性。

表 6-42 两种方法检查室壁收缩运动情况

甲法测定结果	例 数			
	* : 正常	减 弱	异 常	合 计
正常	60	3	2	65
减弱	0	42	9	51
异常	8	9	17	34
合计	68	54	28	150

注: * 代表乙法测定结果。

【实例辨析】 本资料属于双向有序且属性相同的 $R \times C$ 列联表资料。根据研究目的应采用 Kappa 检验对两种方法测定结果的一致性进行统计分析。SAS 运行程序如下, 程序见 nkkysj6_38。

```
data nkkysj6_38;
do r=1 to 3;
do c=1 to 3;
input f @@;
output;
end;
end;
datalines;
60 3 2
0 42 9
8 9 17
;
```

```
run;
proc freq;
weight f;
tables r*c/agree;
test wtkappa;
run;
```

SAS 运行结果：

“r * c”表的统计量	
对称性检验	
统计量(S)	6.6000
自由度	3
Pr > S	0.0858
简单 Kappa 系数	
Kappa	0.6765
渐近标准误差	0.0496
95% 置信下限	0.5792
95% 置信上限	0.7738
加权的 Kappa 系数	
加权的 Kappa	0.6677
渐近标准误差	0.0540
95% 置信下限	0.5618
95% 置信上限	0.7736
H0 检验: 加权的 Kappa = 0	
H0 下的渐近标准误差	0.0641
Z	10.4233
单侧 Pr > Z	< .0001
双侧 Pr > Z	< .0001
样本大小 = 150	

SAS 结果解释：在检验结果中，第一部分是对称性检验的结果， $S = 6.6000$ ， $P = 0.0858 > 0.05$ ，说明此 $R \times C$ 列联表的频数满足对称性假设，即此表中的各频数是关于主对角线对称的。Kappa 系数部分给出了简单和加权 Kappa 统计量的值、渐近标准误、总体 Kappa 值的 95% 可信区间。此处采用加权 Kappa 检验较为合适(可充分利用非对角线上的信息)。加权的 Kappa 系数为 0.6677， $Z = 10.4233$ ， $P < 0.0001$ ，有统计学差异，尚不能认为两种方法的测定结果没有一致性。

10. 高维列联表资料的统计分析

高维列联表资料是指表中所涉及的定性变量的个数 ≥ 3 。根据结果变量的性质不同可分为 3 类，不同的高维列联表资料所用的统计分析方法不同。

- (1) 结果变量为二值变量的高维列联表资料，通常选用对数线性模型或 logistic 回归分析。
- (2) 结果变量为多值有序变量的高维列联表资料，通常选用 CMH 校正的秩和检验或有序变量的累积多重 logistic 回归分析等。
- (3) 结果变量为多值名义变量的高维列联表资料，通常选用 CMH 校正的 χ^2 检验或多值名义变量的扩展多重 logistic 回归分析等。

【例 6-39】 对某地区居民慢性肾脏病 (CKD) 患病情况进行调查，结果见表 6-43。试对此表中的资料进行统计分析。

表 6-43 慢性肾脏病 (CKD) 分期的观测结果

地 区	性 别	调 查 人 数	例 数			
			* : 1 期	2 期	3 期	4 期
城镇	男	2204	63	67	56	19
	女	2214	67	64	78	22
	合计	4418	131	134	41	175
农村	男	2288	86	52	60	24
	女	2251	58	49	62	39
	合计	4539	144	101	122	63

注：* 代表 CKD 分期。

【实例辨析】 表 6-43 中含有 3 个定性变量，分别为地区、性别和 CKD 分期；结果变量为 CKD 分期 (有序变量)，故该表称为结果变量为多值有序变量的三维列联表资料。对其进行统计分析时可选用累计多重 logistic 回归分析。SAS 运行程序如下，程序见 nkkysj6_39。

```
data nkkysj6_39;
do a=1 to 2;
do b=1 to 2;
do c=1 to 4;
input f @@;
output;
end;
end;
end;
datalines;
63 67 56 19
67 64 78 22
86 52 60 24
58 49 62 39
;
proc logistic;
weight f;
model c=a b/selection = stepwise
sle=0.5 sls=0.05;
run;
```

SAS 运行结果：

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept 1	1	-0.2763	0.1972	1.9632	0.1612
Intercept 2	1	0.8418	0.1991	17.8743	<.0001
Intercept 3	1	2.5016	0.2177	132.0539	<.0001
b	1	-0.3318	0.1231	7.2658	0.0070
Odds Ratio Estimates					
Effect	Point Estimate	95% Wald Confidence Limits			
b	0.718	0.564	0.913		

SAS 结果解释：第一部分结果表明，自变量 b (性别) 所对应的统计量具有统计学意义，Wald Chi-Square 值为 7.2658， $P = 0.0070$ ，即性别对慢性肾脏病 (CKD) 的分期有影响。从第二部分结果可知，性别的 $OR = 0.718$ ，女性的分期是男性的 0.718 倍，即女性的分期较男性低。

本节小结

选择合适的统计分析方法对资料进行处理时，需对研究目的、资料类型、试验设计类型、样本含量的大小及各种统计分析方法的应用条件等诸多要素进行综合考虑。只有选择合适的方法，方可真实揭示资料内部的变化规律。

6.4 如何合理应用相关和回归分析

研究在专业上有一定联系的两定量变量之间是否存在直线关系以及如何求得直线回归方程等问题，需要进行直线相关和回归分析。直线相关分析是描述两变量间是否有直线相关关系及相关的方向和密切程度等的方法。但是某些科学研究者在应用相关和回归分析资料时，出现许多错误，如有些研究者不绘制散点图，不管变量间在专业上是否存在一定的联系就直接进行相关和回归分析。也有些研究者将统计学上有意义但专业意义不大的资料也做相关与回归分析等。以下针对内科学杂志中关于相关和回归分析方法错误应用的一些实例，简单介绍如何合理应用简单线性相关和回归分析方法。

6.4.1 散点图呈直线趋势，方可做直线相关和回归分析

【例 6-40】 原文题目：双歧杆菌分泌型粘附素对肠上皮细胞应激反应的影响。目的：观察双歧杆菌分泌型粘附素对肠上皮细胞应激反应后核因子 NF-κBDNA 结合活性的表达和 IL-8 细胞因子表达的影响。统计学处理：所测数据用 SPSS 13.0 统计软件进行 Pearson 相关分析和回归分析。结论：NF-κBDNA 结合活性与 IL-8 mRNA 的表达水平均呈正相关。回归方程为： $Y = 0.0252X + 0.2745$ 。

【实例辨析】 在做直线相关和回归分析性时，首先要绘制散点图用以了解两定量变量之间的直线关系的密切程度和方向。若散点图未显示出明显的线性趋势，则不适合做直线相关分析；如果存在曲线变化趋势，则应对两变量间进行曲线相关分析。

在绘制散点图时，将 (x, y) 的 n 对数值绘在直角坐标系内，得到 n 个点，如果这 n 个点在不平行也不垂直于 x 轴的一条不太宽的长条带内随机地分布着，且不存在着明显的曲线趋势，才可以进一步做直线相关分析。就本资料而言，通过图 6-12 中的散点分布图可以看出，两变量间有明显的线性趋势，各点随机地分布着，故适合做直线相关和回归分析。

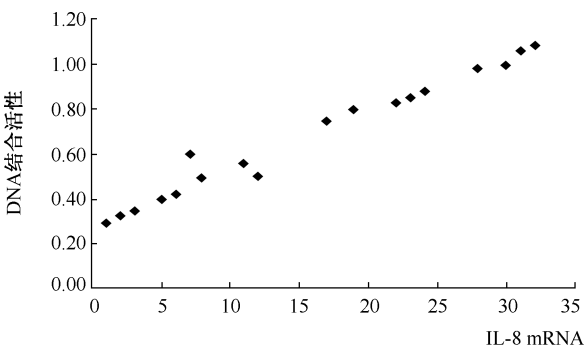


图 6-12 NF-κBDNA 结合活性与细胞因子 mRNA 表达的关系

6.4.2 对关联性强度定位错误

【例 6-41】 原文题目：血清雌激素水平与结肠黏膜组织错配修复基因表达水平的关系。目的：研究体内雌激素对结肠黏膜组织错配修复基因表达的影响。方法：共选取结肠镜检查正常的健康人 42 名。于结肠镜检查前 0.5h 抽取静脉血测定血清雌二醇(E2)水平，结肠镜下于右半结肠取肉眼观正常的结肠黏膜组织，以半定量逆转录(RT)-聚合酶链反应(PCR)和免疫组化染色法评定 DNA 错配修复基因 hMLH1、hMSH2 的表达量，分析血清 E2 水平与结肠黏膜 hMLH1 和 hMSH2 表达量的关系。原文作者在统计学处理中提到“使用 Spearman 秩相关评估各组中 E2 水平与结肠黏膜 hMLH1、hMSH2 表达量及与 PCNA 表达的关系。 $P < 0.05$ 为差异有统计学意义。Spearman r 值 > 0.4 或 < -0.4 为有强关联。

【实例辨析】 本例原作者在血清雌激素水平与结肠黏膜组织错配修复基因表达水平的关系研究中，采用 Spearman 秩相关评估各组中 E2 水平与结肠黏膜 hMLH1、hMSH2 表达量及与 PCNA 表达水平间的关系，并在统计学处理中提及“Spearman 秩相关系数 r 值 > 0.4 或 < -0.4 为有强关联”。这种说法是不够严谨的。一般情况下，相关系数的绝对值越大，则两变量间的关联强度就越大，但是当总体上绝对值大于 0.7 时，才定位为强相关。此处，原作者将 Spearman 秩相关系数绝对值大于 0.4 界定为强相关，是不妥当的。

6.4.3 滥用直线相关分析得出没有实际意义的结论

【例 6-42】 某研究者欲探讨乙型肝炎慢加急性肝功能衰竭(ACLF)患者体内免疫功能抑制与疾病严重程度的关系。收集乙型肝炎 ACLF 患者 27 例(ACIF 组)，活动性慢性乙型肝炎患者 28 例(CHB 组)和健康志愿者 8 名(对照组)的临床资料及外周血标本。以 APACHE III 评分及肝性脑病程度作为反映疾病严重程度的量化指标。统计分析方法：采用 Spearman 双侧检验进行相关性分析。结论：单核细胞 HLA-DR 平均荧光强度与 APACHE III 评分存在负相关性。 $(R^2 = 0.2667, P = 0.0167)$ ，患者 APACHE III 评分越高，单核细胞 HLA-DR 平均荧光强度越低，表明单核细胞 HLA-DR 表达量越低。

【实例辨析】 在考察两变量或多变量之间是否有相关关系时，首先要看这几个变量在专业上是否有联系，是否有价值做相关或回归分析。可结合散点图，考察两变量是否确有直线关系，如果从散点图上看两变量之间具有某种曲线关系，可以试着做曲线拟合，以便更好地揭示它们之间的本质联系。当计算出相关系数并加上假设检验得到有统计学意义的结论后，计算决定系数，并根据决定系数的大小判断两变量之间关系的实际意义。

本例采用 Spearman 秩相关分析(双侧检验)，结论为单核细胞 HLA-DR 平均荧光强度与 APACHE III 评分存在负相关性 $(R^2 = 0.2667, P = 0.0167)$ 。从该结论可知，其决定系数 R^2 仅为 0.2667，由此可知单核细胞 HLA-DR 平均荧光强度与 APACHE III 评分间的关系实际意义并不大。

6.4.4 误用 t 检验方法说明相关回归问题

【例 6-43】 原文题目：胃癌及术后随访患者血清血管生成素-2 含量的临床意义。目的：胃癌患者血清 Ang-2 含量变化与胃癌病理特征的关系。统计学处理：应用 SPSS 12.0 统计软件处理，检测数据以 $\bar{x} \pm s$ 表示，组间数据比较采用 t 检验，率的比较采用 χ^2 检验，相关分析采用 Spearman 等级相关法。 $P < 0.05$ 为差异有统计学意义。结果：血管生成素-2 含量变化与肿瘤分化程度、TNM 病理分期、有无淋巴结转移、有无远处组织转移、浸润深度和肿瘤大小均相关(P 值均 < 0.01)，而与组织学分型、肿瘤部位无相关性(P 值均 > 0.05)，见表 6-44。

表 6-44 胃癌组血清血管生成素-2 含量变化与肿瘤病理特征的关系

肿瘤病理特征	例 数	血管生成素-2(μg/L)	统 计 值	P 值
组织学分型			$r = 0.232$	> 0.05
管状腺癌	58	324.4 ± 46.3		
乳头状腺癌	45	331.6 ± 39.2		
黏液腺癌	31	339.2 ± 33.5		
印戒细胞癌	24	348.9 ± 50.8		
分化程度			$t = 6.208$	< 0.01
高、中分化	63	293.5 ± 41.6		
低、未分化	95	341.3 ± 50.9		
淋巴结转移			$t = 4.841$	< 0.01
有	107	362.7 ± 58.1		
无	51	317.5 ± 47.2		
远处组织转移			$t = 4.860$	< 0.01
有	25	377.2 ± 53.6		
无	133	313.9 ± 60.8		
TNM 分期			$t = 5.854$	< 0.01
I 期和 II 期	59	308.1 ± 41.5		
III 期和 IV 期	99	350.6 ± 45.7		
浸润深度			$t = 3.641$	< 0.01
浆膜内	68	321.6 ± 51.7		
浆膜外	90	354.3 ± 58.9		
肿瘤直径			$t = 10.705$	< 0.01
≤5cm	105	319.7 ± 34.5		
≥5cm	53	362.2 ± 63.6		
肿瘤部位			$r = 0.203$	> 0.05
胃窦部	78	341.2 ± 40.8		
胃底贲门部	35	337.1 ± 35.5		
胃体部	24	331.7 ± 45.7		
广泛区	21	354.9 ± 51.4		

【实例辨析】 由表 6-44 可知，原作者通过检验结果显示胃癌组血清血管生成素-2 含量的变化与肿瘤病理特征(淋巴结转移、远处组织转移、浸润深度、TNM 病理分期和肿瘤大小)均相关(P 值均 < 0.01)。这种分析方法是错误的，因为检验不能用来分析变量间的相关关系，而应采用 Spearman 秩相关分析。对于该资料，根据研究者的研究目的，也可选用多重线性回归分析方法。

6.4.5 误用直线相关分析代替 Spearman 等级相关分析

【例 6-44】 原文题目：环氧合酶-2 与 5-脂氧合酶在溃疡性结肠炎(UC)黏膜中的表达及相互关系。方法：取 32 例 UC 患者结肠黏膜标本根据内镜及组织学进行分级,同时收集健康对照者结肠黏膜标本 26 例。采用实时荧光定量反转录聚合酶链反应(RT-PCR)和免疫组化法检测溃疡性结肠炎患者结肠黏膜 COX-2、5-LOX 的 mRNA 和蛋白表达及其相互关系，见表 6-45。统计学处理：各指标间的相关性分析采用 Pearson 相关性检验。结论：UC 组 COX-2、5-LOX 蛋白阳性表达率与内镜分级相关($P < 0.05$)，即随内镜分级增加的增高其表达也逐渐增高。

表 6-45 COX-2、5-LOX 在 UC 患者结肠黏膜不同内镜分级中的表达率($\bar{x} \pm s$)

分 组	例 数	环氧合酶-2	5-脂氧合酶
对照组	26	10.24 ± 1.44	8.92 ± 1.36
内镜 1 级	10	15.63 ± 1.95	14.71 ± 1.86
内镜 2 级	19	27.71 ± 3.08	22.83 ± 2.97
内镜 3 级	3	40.12 ± 4.11	36.84 ± 3.58

6.4.6 误用卡方检验来说明相关回归才能回答的问题

【例 6-45】 原文题目：Transgelin 在伴随糖尿病的胰腺癌中的表达及其对 SW1990 细胞运动侵袭的影响。目的：探讨骨架相关蛋白 Transgelin 在伴或不伴糖尿病的胰腺癌组织中的表达及其对胰腺癌 SW1990 细胞运动侵袭的影响。方法：根据 92 例胰腺癌患者年龄、性别、是否伴随糖尿病、部位、分化程度、TNM 分期、是否有淋巴结转移、门静脉及腹腔神经受侵犯，统计 Transgelin 免疫组织化学的阳性表达与胰腺癌临床病理特征的关系。结果：除了与糖尿病有关外，Transgelin 蛋白的表达还与淋巴结转移($\chi^2 = 5.778, P = 0.016$)及 TNM 分期密切相关($\chi^2 = 5.966, P = 0.015$)；与患者年龄、性别、肿瘤发生部位、分化程度、门静脉及腹腔神经受侵袭无关($P > 0.05$)，见表 6-46。

表 6-46 Transgelin 的表达与胰腺癌临床病理参数的关系(例)

临床病理参数	总 例 数	Transgelin 表达		χ^2	P 值
		阳 性	阴 性		
年龄				0.199	0.656
<55 岁	54	36	18		
≥55 岁	38	27	11		
性别				0.304	0.581
男	63	42	21		
女	29	21	8		
糖尿病				10.403	0.001
有	45	38	7		
无	47	25	22		
部位				0.358	0.836
胰头	58	41	17		
胰体	14	9	5		
胰尾	20	13	7		
分化程度				2.668	0.102
高 ~ 中分化	52	32	20		
低分化	40	31	9		
淋巴结转移				5.778	0.016
有	39	32	7		
无	53	31	22		
门静脉或腹腔神经受侵袭				3.443	0.064
有	44	26	18		
无	48	37	11		
TNM 分期				5.966	0.015
I + II	37	20	17		
III + IV	55	43	12		

【实例辨析】 本例采用一般卡方检验的结果来说明 Transgelin 蛋白的表达与淋巴结转移及 TNM 分期密切相关，这是不正确的。一般卡方检验只能用来考察列联表资料的行变量和列变量之间是否独立，并不能说明行变量和列变量之间存在相关关系。本例要说明 Transgelin 蛋白的表达水平与患者临床病理参数间的关系，可采用 Spearman 秩相关分析、多因素 logistic 回归分析等。

【例 6-46】 原文题目：模型基质金属蛋白酶-2 在食管癌中的表达及其临床意义。作者对某医院收住的 103 例食管癌患者的 MT2-MMP 表达与其病理特征的关系进行相关性分析。统计学处理：运用 GraphPad Prism 4.0 统计软件对数据进行 t 检验, χ^2 检验及生存分析、Log-rank 检验, $P < 0.05$ 为差异有统计学意义。结论：MT2-MMP 表达与患者的肿瘤大小相关($P = 0.044$)，而与其他临床病理特征无关(P 值均 >0.05)，见表 6-47。

表 6-47 模型基质金属蛋白酶-2 在食管癌中的表达水平与患者临床病理特征的关系

临床病理特征	例 数	模型基质金属蛋白酶-2 表达水平				P 值
		H-score \leq 105 (例)	%	H-score $>$ 105 (例)	%	
性别						0.2994
男	75	46	61.3	29	38.7	
女	28	14	50.0	14	50.0	
年龄						0.4861
<60 岁	63	35	55.6	28	44.4	
\geq 60 岁	40	25	62.5	15	37.5	
肿瘤大小						0.0440
<3.5cm	38	27	71.1	11	28.9	
\geq 3.5cm	65	33	50.8	32	49.2	
肿瘤分级						0.7277
T_1	13	8	61.5	5	38.5	
T_2	46	27	58.7	19	41.3	
T_3	33	19	57.6	14	42.4	
T_4	11	6	54.5	5	45.5	
淋巴结转移						0.1563
无	54	35	62.1	33	37.9	
有	49	25	37.5	10	62.5	
远处转移						0.0670
无	87	54	62.1	33	37.9	
有	16	6	37.5	10	62.5	
TNM 分期						0.2128
I	13	8	61.5	5	38.5	
II	56	34	60.7	22	39.3	
III	18	12	66.7	6	33.3	
IV	16	6	37.5	10	62.5	

【实例辨析】 本例所犯错误在于用一般卡方检验来说明基质金属蛋白酶-2 在食管癌中的表达水平与患者临床病理特征间的关系，这是不正确的。要说明基质金属蛋白酶-2 的表达水平与患者多个临床病理特征间的关系，则可以模型基质金属蛋白酶-2 表达水平为因变量，以病理特征为自变量进行二值 logistic 回归分析。另外，就本例资料而言，还可以 H-score 评分(假设服从

正态分布)的具体数值为因变量,以性别、年龄及肿瘤大小等病理特征为自变量进行多重线性回归分析。

由此可知,在研究多个自变量与一个因变量的关系时,可选用多因素 logistic 回归分析或多重线性回归分析。具体选用何种方法主要根据因变量的类型而定。当因变量的取值为定量资料且服从正态分布时,一般选用多重线性回归分析;当因变量的取值为定性的二值或多值变量时,则可以采用 logistic 回归分析。

6.4.7 多重线性回归分析的合理应用

多重线性回归分析自变量的赋值策略:在进行多重线性回归分析时,自变量赋值问题需要首要考虑。其赋值方法与多因素 logistic 回归分析自变量赋值方法相似,均需根据自变量类型选择合适的赋值方法。当自变量类型为多值名义变量时,可采用哑变量赋值方法;当自变量为多值有序变量时,具体赋值方法主要有三种:

- (1)当根据专业知识确定有序变量每上升或下降一个等级,因变量的变化幅度都是相等时,可以给有序变量的各等级赋值“1, 2, 3, …”。
- (2)当根据专业知识确定有序变量每上升或下降一个等级、因变量的变化幅度不等时,可以结合专业知识给有序变量的各等级赋予相应的数值。
- (3)当通过专业知识无法判断有序变量的等级改变会对因变量的改变有何影响时,应把“多值有序变量”视为“多值名义变量”,采用哑变量赋值方法。

本节小结

- (1)相关关系是一种共变关系,有相关关系表达的观察指标之间不一定是因果关系,也不一定有内在联系。因此,不能单纯依靠相关系数的统计学检验的结果“证明”相关或因果关系的存在。
- (2)无论是进行相关分析还是回归分析,首先都应绘制散点图,以判断两变量间的关系是否存在某种直线或曲线变化趋势,有无离群点等。
- (3)进行直线回归分析要有意义,任何 n 对数据都能求出相关系数和回归方程,不能因此就说明这两种事物间存在内在联系,更不能立即确定有因果关系,而可能是一种伴随关系。
- (4)相关或回归分析有无实用价值,不能仅凭假设检验的 P 值来判断相关或回归效果的优劣,应结合决定系数 r^2 数值的大小($r^2 > 0.5$)和具体专业知识对相关分析有无实用价值做出科学评价;结合专业知识和残差分析对回归方程有无实用价值做出科学评价。

6.5 如何合理应用诊断试验评价指标

诊断试验在临床上的应用涉及面甚广,包括疗效的判断、药物毒副作用监测、疾病预后的判断,以及应用于普查、筛检无症状的病人等。用于不同场合的诊断试验有不同的要求,不同的诊断试验本身又有一定的特性,临床医师需要对诊断试验进行科学的评价。科学地评价诊断试验是临床医师选择诊断试验的基础。

评价诊断试验的优劣必须以金标准(gold standard)作为参照,没有金标准的诊断试验评价是没有科学性的。所谓金标准,是指当前临床医学界所公认的诊断某种疾病最可靠的方法。利用金标准能正确地区分某人是“有病”还是“无病”。临床诊断常用的金标准包括病理学诊断(组织活检和尸检)、外科手术发现、特殊的影像学诊断等。金标准的选择至关重要,如果采用的金标准选择不当,就会造成分类错误,从而影响诊断试验的正确评价。

6.5.1 常用诊断试验评价指标

用某一诊断试验和金标准对诊断对象进行判别，无外乎会出现 4 种情况，一般将其整理成如表 6-48 所示的形式，其本质上为一个配对四格表。

表 6-48 诊断试验四格表

某诊断试验结果	例 数		
	金标准：有病 D +	无病 D -	合 计
阳性 T +	a(真阳性)	b(假阳性)	a + b
阴性 T -	c(假阴性)	d(真阴性)	c + d
合计	a + c	b + d	N(a + b + c + d)

1. 常用评价指标

评价诊断试验的常用指标有灵敏度、特异度、误诊率、漏诊率。

(1) 灵敏度(sensitivity, S_e): 又称真阳性率, 是实际患病且被试验诊断为患者的概率, 也就是患者被诊断为阳性的概率; 反映诊断试验将实际有病的人正确地判为患者的检出能力。

$$\text{灵敏度} = \frac{a}{a + c} \times 100\%$$

(2) 特异度(specificity, S_p): 又称真阴性率, 是实际未患病而被试验诊断为非患者的概率, 也就是非患者被诊断为阴性的概率。反映诊断试验鉴别非患者的能力。

$$\text{特异度} = \frac{d}{b + d} \times 100\%$$

(3) 误诊率: 又称假阳性率(α), 指实际未患病而被诊断试验诊断为患病的概率; 反映非患者被错误诊断的可能性。

$$\text{假阳性率} = \frac{b}{b + d} \times 100\%$$

(4) 漏诊率: 又称假阴性率(β), 指实际患病但被试验诊断为非患病的概率, 反映患者被遗漏诊断的可能性。

$$\text{假阴性率} = \frac{c}{a + c} \times 100\%$$

2. 预测评价指标

评价诊断试验的预测的准确性常用阳性预测值和阴性预测值两个指标。

(1) 阳性预测值(positive predictive value, PV_+)是指由试验诊断为阳性者, 确为患者的概率。

$$\text{阳性预测值} = \frac{a}{a + b} \times 100\%$$

(2) 阴性预测值(negative predictive value, PV_-)是指由试验诊断为阴性者, 确为非患者的概率。

$$\text{阴性预测值} = \frac{d}{c + d} \times 100\%$$

3. 综合评价指标

单独用灵敏度和特异度评价诊断试验, 可能会出现一个诊断试验灵敏度高, 而另一个诊断试验特异度高, 无法判断哪一个诊断试验更好的现象。所以有必要将灵敏度和特异度结合起来运用综合评价指标来评价诊断试验优劣。综合评价指标主要有正确率、Youden 指数、比数积、阳性似然比、阴性似然比等。

(1) 正确率(π)：又称总符合率，表示观察结果与实际结果的符合程度，反映了正确诊断患者与非患者的综合能力。实际上是灵敏度和特异度的加权均数。

$$\text{正确率} = \frac{a + d}{N} \times 100\%$$

(2) Youden 指数(YI)：反映了诊断试验真实性的综合指标。范围介于 -1 ~ 1 之间，其值越大，说明诊断试验真实性越好。

$$YI = \text{灵敏度} + \text{特异度} - 1$$

(3) 比数积(OP)：表示患者中诊断阳性数、阴性数之比与非患者中诊断阴性数、阳性数之比的乘积。其值越大，诊断价值越高。

$$\text{比数积} = \frac{\text{敏度}}{1 - \text{敏度}} \cdot \frac{\text{特异度}}{1 - \text{特异度}} = \frac{ad}{bc}$$

(4) 阳性似然比(LR₊)：真阳性率与假阳性率之比。

$$\text{阳性似然比} = \frac{a}{(a + c)} \bigg/ \frac{b}{(b + d)}$$

(5) 阴性似然比(LR₋)：假阴性率与真阴性率之比。

$$\text{性似然比} = \frac{c}{(a + c)} \bigg/ \frac{d}{(b + d)}$$

6.5.2 诊断试验的正确应用

(1) 诊断试验必须以金标准为对照，没有金标准的诊断试验评价是没有意义的。

(2) 诊断试验中应采用盲法，避免主观因素对结果的干扰。

(3) 用于诊断试验评价的研究对象应具有代表性，研究对象应包括各种类型病例，以提高诊断试验的普遍性及临床推广价值。

(4) 在评价诊断试验时，要说明研究对象所包括的病例和对照的来源。因为不同来源病人(如三级医院和乡镇医院)由于来源不同，诊断水平存在较大差异；交代清楚诊断试验研究对象的来源，有利于专业人员对诊断试验有一个客观的评价和认识。

(5) 需要顾及诊断试验的实用性。评价诊断试验除了考虑上述的评价指标外，还需要考虑诊断试验是否方便、快捷、廉价、有无副作用和危害性，是否符合伦理学要求等方面。

6.5.3 诊断试验常见错误及错误辨析

纵观以往发表的诊断试验研究论文，发现不少存在设计不合理，评价指标使用不当等问题，还有不少诊断试验评价论文仅采用 *t* 检验或 χ^2 检验比较某诊断试验与金标准诊断结果，就得出某诊断试验优于金标准，值得广泛推广应用的结论。常见的错误归纳总结如下。

(1) 金标准选择缺陷：无金标准或以不可靠的诊断方法作为金标准。选择金标准是整个诊断试验的核心内容，如无金标准或金标准选择不当，则整个试验的准确度就无从谈起。

例如有篇报道拟以妇产科门诊患有不同程度的宫颈感染或不孕症患者共 77 例为研究对象，评价 PCR 技术检测患者宫颈管细胞碎片中沙眼衣原体的敏感度。结果：免疫荧光技术阳性 7 例，且这 7 例阳性标本的 PCR 检测结果也为阳性，据此认为 PCR 方法敏感度为 100%。解析：诊断衣原体公认、可靠的金标准方法是及时从细胞培养中分离衣原体，该文仅以免疫荧光技术检测结果为标准，据此说明 PCR 的敏感度为 100% 是不正确的，应该选择细胞培养分离方法作为金标准进行比较研究，而不应该以免疫荧光技术作为金标准。

(2) 研究对象不包括易混淆病例，不具有代表性。

研究对象应该包含容易混淆的病例，保证研究对象具有较好的代表性，保证评价结果具有较好的外推性。例如研究某种诊断试验队肺癌患者诊断效果，研究对象应该选择肺癌患者、健康人，以及容易与肺癌患者混淆的肺结核、肺炎性假瘤、纵隔肿瘤及肺部良性肿瘤等患者，而不应该仅选择肺心病患者。

(3) 没有明确交代研究对象的选取方法和来源。

(4) 没有交代研究对象的疾病程度。

研究对象最好包含轻、中、重各种类型的病例，而且应该对研究对象疾病程度有详尽的交代。

(5) 没有设置对照组。

(6) 诊断试验中未贯彻盲法，无法避免主观因素对结果的干扰。

若把试验结果作为确立诊断的临床依据的一部分，或者把最后的诊断或其他临床信息作为得出试验结果的考虑依据。可能因为临床提示，增加了诊断性试验结果和金标准结果的一致性，从而夸大了其准确性和使用的价值。因此试验的评价应独立进行，同时与金标准的比较也应贯穿盲法，以避免因偏倚因素造成诊断性指标的错误或不当的评估。

(7) 评价指标概念混淆，导致指标计算错误。

6.6 如何合理应用 ROC 曲线分析

6.6.1 ROC 曲线简介

受试者工作特征曲线(Receiver operator characteristic curve, ROC 曲线)起源于电子信号观测理论,从 20 世纪 80 年代起广泛应用于医学诊断试验的评价。ROC 曲线将诊断试验结果划分为若干临界点,以每个临界点对应的(1-特异度)为横坐标、灵敏度为纵坐标,作图得到的曲线。ROC 可以分析有序分类资料与连续型资料两种类型资料。

【例 6-47】 为了研究肌酸激酶(CK)诊断心肌梗塞的作用,对金标准诊断为心肌梗塞的 230 例病人和 130 名正常人分别测定了每个人的 CK 值,如表 6-49 所示。试绘制其 ROC 曲线。

表 6-49 有序分类 ROC 分析资料

CK 值	例 数		
	金标准：病人	正 常 人	合 计
1 ~	2	88	90
40 ~	13	26	39
80 ~	118	15	133
280 ~	97	1	98
合计	230	130	360

对于这种 4 级分类资料,按级别从大到小排列,以前 3 个分类(即不考虑最小的分类点)作为诊断界点,大于等于诊断界点者为阳性,小于该值者为阴性。这样,可整理出 3 个四格表,即表 6-50 ~ 表 6-52(纵向的标目都是“金标准”,从略)。

表 6-50 诊断界点 =40

CK 值	例 数	
	病 人 组	正常人组
<40	2	88
≥40	228	42

表 6-51 诊断界点 =80

CK 值	例 数	
	病 人 组	正常人组
<80	15	114
≥80	215	16

表 6-52 诊断界点 =280

CK 值	例 数	
	病 人 组	正常人组
<280	113	129
≥280	97	1

然后分别按照灵敏度和特异度计算公式，算出各种诊断界值的灵敏度(S_e)和特异度(S_p)，如表 6-53 所示。

表 6-53 4 种诊断方法的灵敏度和特异度

诊断界值	S_e	S_p	$1 - S_p$
1	1	0	1
40	0.9913	0.6769	0.3231
80	0.9348	0.8769	0.1231
280	0.4127	0.9923	0.0077

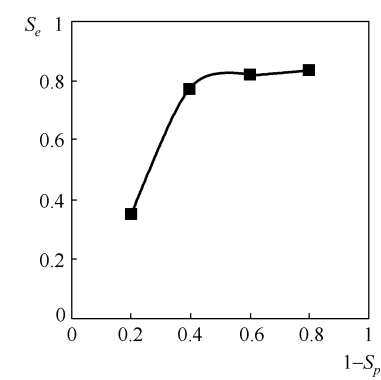


图 6-13 ROC 曲线

对表 6-53 中的数据，以假阳性率(1-特异度)为横轴，以灵敏度为纵轴，在算术坐标纸上标出 ROC 曲线工作点，然后用直线连接各相邻两点，就得到了 ROC 曲线，如图 6-13 所示。

【例 6-48】 假设某诊断试验的病例组和对照组分别有 5 个和 4 个受试者，其检测结果见表 6-54。试计算所有可能的 TPR 和 FPR 值(显然，样本含量太少，这里只是为了便于叙述)。

将这 9 个数据从大到小排列，以前 8 个数(不考虑最小值 1.7)分别作为诊断界点，大于等于诊断界点者判为阳性，小于该值者判为阴性。这样，可整理成 8 个四格表(纵向的标目都是“金标准”，从略)。

表 6-54 连续性 ROC 分析资料

金 标 准		诊 断 结 果				
病例组	16.5	13.5	12.8	11.2	5.0	
对照组	8.5	6.4	4.6	1.7		

诊断界点 = 16.5

诊断结果	例数	病	例	对	照
+	1	0			
-	4	4			

诊断界点 = 13.5

诊断结果	例数	病	例	对	照
+	2	0			
-	3	4			

诊断界点 = 12.8

诊断结果	例数	病	例	对	照
+	3	0			
-	2	4			

诊断界点 = 11.2

诊断结果	例数	病	例	对	照
+	4	0			
-	1	4			

诊断界点 = 8.5

诊断结果	例数	病	例	对	照
+	4	1			
-	1	3			

诊断界点 = 6.4

诊断结果	例数	病	例	对	照
+	4	2			
-	1	2			

诊断界点 = 5.0

诊断结果	例数	病	例	对	照
+	5	2			
-	0	2			

诊断界点 = 4.6

诊断结果	例数	病	例	对	照
+	5	3			
-	0	1			

然后分别按照灵敏度和特异度计算公式，算出 8 个诊断界值的灵敏度(S_e)和特异度(S_p)，以灵敏度为纵坐标，(1-特异度)为横坐标，标出 ROC 曲线工作点，然后用直线连接各相邻两点构建 ROC 曲线即可。

6.6.2 ROC 曲线下面积计算

ROC 曲线下面积(记为 A_z)因其不受患病率和诊断界值的影响以及可对两个诊断试验的准确度进行综合比较，因而是一种全面、准确评价诊断试验的有效工具。ROC 曲线与对角线间的面积

反映了诊断试验的准确性，面积越大，准确性越高；成为目前公认的标准评价指标。同时，ROC 曲线也被用于确定检测的最佳诊断界值；通常选择曲线上尽量靠近左上方的点，并结合专业情况，确定最佳临界值，从而使试验的灵敏度和特异度均较高。

ROC 曲线下面积的取值范围为 0.5~1.0。完全无价值的诊断 $A_z = 0.5$ ，完全理想的诊断 $A_z = 1$ 。一般认为对于一个诊断试验，ROC 曲线下面积在 0.5~0.7 之间时诊断价值较低，在 0.7~0.9 之间时诊断价值中等，在 0.9 以上时诊断价值较高。

ROC 曲线下面积的估计有参数法和非参数法，均适用于连续性资料或有序资料的诊断试验的准确度评价。

(1)非参数法：根据实验结果直接计算绘制 ROC 曲线所需的工作点（即灵敏度，1-特异度），由此绘制的 ROC 曲线称为经验 ROC 曲线，其曲线下面积可由梯形规则计算得到。非参数法因其没有限制条件，适用于任何诊断试验 ROC 曲线下面积的估计；可通过 SPSS、SAS 等统计软件实现。

(2)参数法：假设患者和非患者的实验结果均符合正态分布，根据试验结果拟合双正态模型，由模型拟合的 ROC 曲线称为拟合 ROC 曲线或光滑 ROC 曲线。参数法估计 ROC 曲线下面积可通过软件 ROCKT0.9 β 来实现。参数法缺点是当样本含量较小时，或试验检测值及其变量变换值均远离参数法应用条件时，结果可能会严重偏离真实值。

得出的 ROC 曲线下面积是否与完全无价值的 $A_z = 0.5$ 有统计学差异，可近似采用标准正态离差 $Z = \frac{A_z - 0.5}{S_e(A_z)}$ 进行假设检验。根据 $A_z \pm u_\alpha S_e(A_z)$ 可计算 A_z 的 100(1 - α)% 的置信区间。

如果需要比较两个诊断试验曲线下面积之间的差异是否有统计学意义，当两曲线下面积分别为 A_{z1} 和 A_{z2} ，对应的标准误分别为 $S_e(A_{z1})$ 和 $S_e(A_{z2})$ ，且面积之间的相关系数很小时，可利用公式 $Z = \frac{|A_{z1} - A_{z2}|}{\sqrt{S_e(A_{z1})^2 + S_e(A_{z2})^2}}$ 实现前述目的；当两曲线下面积间的相关系数较大时，则需要考虑其相关性问题的，因公式较复杂请参阅相关文献。

6.6.3 ROC 曲线分析的 SAS 实现

SAS 可以处理二分类结果“金标准”为因变量，每组的测量值（或有序分类值）为解释变量，利用 LOGISTIC 过程，获得每个可能截断点对应的灵敏度、（1-特异度）等指标，可获得非参数法估计的曲线下面积。

(1)对于类似表 6-54 的连续型数据实现 ROC 分析 SAS 程序如下，程序见 nkkysj6_54。

行号	程序语句
01	Data nkkysj6_54 ;
02	Input group \$ num;
03	Do i =1 To num;
04	Input test_val@@;
05	Output;
06	End;
07	Cards;
08	Case 5
09	16.5 13.5 12.8 11.2 5.0
10	Control 4
11	8.5 6.4 4.6 1.7
12	;

```
13 Proc Logistic Data = nkkysj6_54;
14 Model group=test_val/ Scale=none Outroc=roc1;
15 Proc Print;
16 Run;
```

SAS 程序说明：02 行中 group 的 case 表示病例组，control 表示对照组，09 行和 11 行所给数据分别为每组测量值个数，test_val 表示诊断试验测量数据。主要输出结果如下：

Model Fit Statistics					
		Intercept		Intercept	
Criterion		Only		and Covariates	
AIC		14.365		10.905	
SC		14.563		11.299	
-2 Log L 12.365		6.905			
Testing Global Null Hypothesis: BETA = 0					
Test	Chi-Square	DF	Pr > ChiSq		
Likelihood Ratio	5.4607	1	0.0194		
Score	4.4156	1	0.0356		
Wald	2.6937	1	0.1007		
Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard	Wald	Pr > ChiSq
			Error	Chi-Square	
Intercept	1	-3.8912	2.5342	2.3577	0.1247
test_val	1	0.4951	0.3016	2.6937	0.1007
Odds Ratio Estimates					
Effect	Estimate	Point	95% Wald		
		Confidence	Limits		
test_val	1.641		0.908	2.963	
Association of Predicted Probabilities and Observed Responses					
Percent Concordant 90.0		Somers'D	0.800		
Percent Discordant		10.0	Gamma	0.800	
Percent Tied		0.0	Tau-a	0.444	
Pairs		20	c	0.900	

结果中的 c=0.900，相当于非参数法计算的 ROC 曲线下面积。该程序也可输出各种可能截断点对应的灵敏度（即真阳性率）与（1-特异度），结果如下；这正是绘制 ROC 曲线所需要的工作点数据。

序号	概率	真阳性	真阴性	假阳性	假阴性	灵敏度	1-特异度
Obs	_PROB_	_POS_	_NEG_	_FALPOS_	_FALNEG_	_SENSIT_	_1MSPEC
1	0.98631	1	4	0	4	0.2	0.00
2	0.94224	2	4	0	3	0.4	0.00
3	0.92023	3	4	0	2	0.6	0.00

4	0.83935	4	4	0	1	0.8	0.00
5	0.57853	4	3	1	1	0.8	0.25
6	0.32676	4	2	2	1	0.8	0.50
7	0.19530	5	2	2	0	1.0	0.50
8	0.16604	5	1	3	0	1.0	0.75
9	0.04523	5	0	4	0	1.0	1.00

(2)对于类似表 6-49 的有序分类数据实现 ROC 分析 SAS 程序如下，程序见 nkkysj6_49。

```
行号      程序语句
01      Data nkkysj6_49;
02      Do group=1,0;
03      Do test_val=1 TO 4;
04      Input freq@@;
05      Output; End; End;
06      Cards;
07      2 13 118 97
08      88 26 15 1
09      ;
10      Proc Logistic Descending;
11      Model group=test_val/ Scale=none Outroc=rocl;
12      FREQ freq;
13      Proc Print;
14      Run;
```

SAS 程序说明：02 行中的 group = 1 表示病例组，group = 0 表示对照组，分类 1 ~ 4 赋值到变量名 test_val(03 行)。因每一分类数据是频数，所以程序中采用了 FREQ 语句(04 行)。07 和 08 行输入数据。输出主要结果如下：

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	90.7	Somers'D	0.893
Percent Discordant	1.4	Gamma	0.970
Percent Tied	8.0	Tau-a	0.413
Pairs	29900	c	0.946

结果中 c = 0.946，表明 ROC 曲线下面积为 0.946。同时输出 ROC 曲线绘制所需要各工作点数据。

序号	概率	真阳性	真阴性	假阳性	假阴性	灵敏度	1-特异度
Obs	_PROB_	_POS_	_NEG_	_FALPOS_	_FALNEG_	_SENSIT_	_1MSPEC
1	0.99243	97	129	1	133	0.42174	0.00769
2	0.88658	215	114	16	15	0.93478	0.12308
3	0.31783	228	88	42	2	0.99130	0.32308
4	0.02702	230	0	130	0	1.00000	1.00000

(吕军城 石福艳)

第7章 协方差分析

7.1 协变量与协变量控制的基本思想

7.1.1 协变量的概念

方差分析可以进行两组或两组以上均数的比较,处理因素通常要求是定性的且可以控制的。然而,实际工作中有时某些定量的因素对于结果有重要影响,却由于实验设计疏忽、实验条件限制等原因造成这些因素未加控制或难以控制。例如降压药物疗效考核的临床试验中,患者的初始血压对血压的下降量有影响,但患者的初始血压水平难以控制。这时,需要在设计中设法扣除这些定量因素的影响,使得其在定性因素各水平组中达到均衡,协方差分析就是实现前述要求的一种统计分析方法。

线性回归分析中影响 Y 的一般为定量变量,方差分析的影响 Y 的一般为定性因变量,而协方差分析(analysis of covariance)是将线性回归分析与方差分析结合起来的一种统计学分析方法,影响 Y 的既可以是定量变量也可以是定性变量。

协方差分析的基本思想是,将无法加以或未加以控制的定量变量 X 对定量因变量 Y 的影响看成协变量(covariate),建立因变量随协变量变化的线性回归方程,利用回归关系把 X 值化为相等后再进行各组 Y 的修正均数(adjusted means)间差别的假设检验,通过这种方法从总平方和中把 X 对 Y 的回归平方和扣除,对残差平方和做进一步分解后再进行方差分析,以更好地评价各处理的效应。协变量是不受处理影响的连续变量,它的取值在研究前或者研究中可以被观察到。实质上,就是从 Y 的总平方和中扣除协变量 X 对 Y 的回归平方和,对残差平方和做进一步分解后再进行方差分析,以更好地评价各种处理的效应。

7.1.2 协方差分析的适用条件

协方差分析有两个重要的应用条件:一是观测变量相对独立,服从正态分布,且每个因素各水平组的总体方差应满足齐性要求;二是总体客观存在因变量对协变量的线性回归关系且斜率相同,即各样本回归系数 b 有统计学意义且各回归系数间差别无统计学意义。也就是说,只有样本资料进行方差齐性检验和回归系数的假设检验满足条件,或经过变量变换后满足条件才能进行协方差分析。

协方差分析中比较的是修正均数,也就是假设协变量取值固定在其总均数时的观察变量的均数。修正均数间的差别与实际均数间差别并不一样,只是用于合理比较。协方差分析在均数间差别不大时做比较好,所以在协方差分析之前最好先对协变量均数间的差别做假设检验。如果多因素研究中有一个或多个因素(协变量)不能或难以控制,并且这个或这些协变量对观察变量可能有影响,就应该选择多元协方差分析或多元回归分析。

协方差分析法的使用很广泛,协方差分析中的协变量既可以是一个,也可以是多个。协方差分析可用于处理不同实验设计类型下的定量资料,如完全随机设计、随机区组设计(配伍组设计)、拉丁方设计、析因设计等,解决问题的基本思想是相似的。

7.2 协方差分析的方法与步骤

设有 G 组回归样本，每组有 n 对观察值，如表 7-1 所示。

表 7-1 G 组回归样本观察

X_1	Y_1	X_2	Y_2	...	X_G	Y_G
第 1 组		第 2 组		...	第 G 组	
X_{11}	Y_{11}	X_{21}	Y_{21}	...	X_{G1}	Y_{G1}
X_{12}	Y_{12}	X_{22}	Y_{22}	...	X_{G2}	Y_{G2}
X_{13}	Y_{13}	X_{23}	Y_{23}	...	X_{G3}	Y_{G3}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
X_{1n}	Y_{1n}	X_{2n}	Y_{2n}	...	X_{Gn}	Y_{Gn}

首先计算各组的和 $\sum X_i$ 、 $\sum Y_i$ ，平方和 $\sum X_i^2$ 、 $\sum Y_i^2$ ，积和 $\sum X_i Y_i$ ，均数 \bar{X}_i 、 \bar{Y}_i 及其合计项。然后按照下列步骤进行方差分析。

(1) 利用合计项各数据计算校正数 C_1 、 C_2 、 C_3 ，以及总变异的离均差平方和 l_{XX} 、 l_{YY} ，积和 l_{XY} 及自由度 ν 。

$$C_1 = \frac{(\sum X)^2}{N} \tag{7-1}$$

$$C_2 = \frac{(\sum Y)^2}{N} \tag{7-2}$$

$$C_3 = \frac{\sum X \cdot \sum Y}{N} \tag{7-3}$$

总的均差平方和：

$$ll_{XX} = \sum X^2 - C_1 \tag{7-4}$$

$$ll_{YY} = \sum Y^2 - C_2 \tag{7-5}$$

$$ll_{XY} = \sum XY - C_3 \tag{7-6}$$

$$\nu = N - 1$$

(2) 计算各处理组间的离均差平方和 l_{XX} 、 l_{YY} ，积和 l_{XY} 及自由度 ν 。

组间均差平方和：

$$ll_{XX} = \sum \frac{(\sum X_i)^2}{n_i} - C_1 \tag{7-7}$$

$$ll_{YY} = \sum \frac{(\sum Y_i)^2}{n_i} - C_2 \tag{7-8}$$

$$ll_{XX} = \sum \frac{(\sum X_i)(\sum Y_i)}{n_i} - C_3 \tag{7-9}$$

$$\nu = G - 1$$

(3) 列出协方差计算表，将上述第(1) ~ 第(2)步结果列入表的左侧部分，再由总变异的 l_{XX} 、 l_{YY} 、 l_{XY} 及 ν 减去与处理组相应的各值，得到组内的 l_{XX} 、 l_{YY} 、 l_{XY} 及 ν 。

(4) 计算回归估计误差平方和 $\sum (Y - \hat{Y})^2$ 及自由度，其中总的组内平方和分别计算

$$\sum (Y - \hat{Y})^2 = l_{YY} - \frac{l_{XY}^2}{l_{XX}}$$

(7-10)

自由度是左侧相应自由度减 1。组内估计误差均方和修正均数均方分别是修正均数和组内误差平方和除以相应自由度。

(5)求 F 统计量。

$$F = \frac{\text{修正均数均方}}{\text{组内均方}}$$

(7-11)

并以修正均数自由度为第一自由度，以组内自由度为第二自由度查 F 表得到 P 值。按照检验水准 α 得统计推断。若 $P \leq \alpha$ ，则可认为各总体修正均数间总的来说有差别。究竟哪两组间有差别需要进一步做多重比较。

(6)若需计算求各修正均数，可先求得公共回归系数 b_c ，再算出修正均数 Y_i^* 。

$$b_c = \frac{\text{组内 } l_{XY}}{\text{组内 } l_{XX}}$$

(7-12)

$$Y_i^* = \bar{Y}_i - b_c (\bar{X}_i - \bar{X})$$

(7-13)

(7)如果需要，使用 q 检验对修正均数进行多重比较。

$$q = \frac{|\bar{Y}_A - \bar{Y}_B|}{\sqrt{\frac{S^2_{Y \cdot X}}{n_0} \left[1 + \frac{\text{组间 } l_{XX}}{\text{组内 } (a - 1) l_{XX}} \right]}}$$

(7-14)

其中， \bar{Y}_A 、 \bar{Y}_B 为多重比较中任意两个对比组的修正均数， $S^2_{Y \cdot X}$ 为组内估计误差的均方， n_0 为每组平均例数， a 为将修正均数按大小顺序排列后两两对比均数间包含的组数。再根据误差估计自由度和 a ，查 q 界值表得到 P 值，并按照预设的检验水准 α 做出相应统计推断。

单因素多水平设计一元定量资料协方差分析的基础数据模式及其计算模式分别见表 7-2 和表 7-3。

表 7-2 单因素多水平设计一元定量资料协方差分析的基础数据模式

		X_A	Y_A	X_B	Y_B	...	X_G	Y_G		
		第 A 组		第 B 组			第 G 组			
	1	X_{1A}	Y_{1A}	X_{1B}	Y_{1B}	...	X_{1G}	Y_{1G}		
	2	X_{2A}	Y_{2A}	X_{2B}	Y_{2B}	...	X_{2G}	Y_{2G}		
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	合计	
	n	X_{nA}	Y_{nA}	X_{nB}	Y_{nB}	...	X_{nG}	Y_{nG}	X	Y
$\sum X_i$	$\sum Y_i$	$\sum X_A$	$\sum Y_A$	$\sum X_B$	$\sum Y_B$		$\sum X_G$	$\sum Y_G$	$\sum X$	$\sum Y$
$\sum X_i^2$	$\sum Y_i^2$	$\sum X_A^2$	$\sum Y_A^2$	$\sum X_B^2$	$\sum Y_B^2$		$\sum X_G^2$	$\sum Y_G^2$	$\sum X^2$	$\sum Y^2$
\bar{X}_i	\bar{Y}_i	\bar{X}_A	\bar{Y}_A	\bar{X}_B	\bar{Y}_B		\bar{X}_G	\bar{Y}_G	\bar{X}	\bar{Y}
S_{X_i}	S_{Y_i}	S_{X_A}	S_{Y_A}	S_{X_B}	S_{Y_B}		S_{X_G}	S_{Y_G}	S_X	S_Y
$\sum X_i Y_i$		$\sum X_A Y_A$		$\sum X_B Y_B$			$\sum X_G Y_G$		$\sum XY$	
b_i		b_A		b_B			b_G		b_C	
a_i		a_A		a_B			a_G		—	
r_i		r_A		r_B			r_G		—	

【例 7-1】 表 7-4 资料为某地方病研究所测定 8 名正常儿童和 12 名大骨节患儿的年龄与其尿肌酐含量的结果，比较这两种儿童人群的尿肌酐含量是否不同，且同时要扣除年龄的影响。

表 7-3 协方差分析计算模式

变异来源	ν	离均差平方和及积和			估计误差			F
		l_{XX}	l_{XY}	l_{YY}	l_{XX}	l_{XY}	l_{YY}	
总变异								
组间变异								
组内变异								
修正均数								

表 7-4 8 名正常儿童和 12 名大骨节病患儿的年龄 X (岁)与尿肌酐含量 Y (mmol/24h)

		正常儿童		大骨节病患者			
		X	Y	X	Y		
	1	13	3.54	10	3.01		
	2	11	3.01	9	2.83		
	3	9	3.09	11	2.92		
	4	6	2.48	12	3.09		
	5	8	2.56	15	3.98		
	6	10	3.36	16	3.89		
	7	12	3.18	8	2.21		
	8	7	2.65	7	2.39		
	9			10	2.74		
	10			15	3.36		
	11			13	3.54	合计	
	12			11	3.01	X	Y
	n_j	8		12		20	
$\sum X_j$	$\sum Y_j$	76	23.87	137	36.97	213	60.84
$\sum X_j^2$	$\sum Y_j^2$	764	72.27	1655	117.10	2419	189.37
$\sum X_j Y_j$		232.61		438.05		670.66	
\bar{X}_j	\bar{Y}_j	9.50	2.98	11.42	3.08	10.46	3.03

成组设计一元定量资料的协方差分析，以儿童的年龄为协变量 X 做协方差分析。
 H_0 : 两种儿童人群尿肌酐含量修正均数相等。
 H_1 : 两种儿童人群尿肌酐含量修正均数不等。
 $\alpha = 0.05$ 。
进行初步计算(表 7-4 下半部分)，并运算出相应的校正数的总的、组间和组内里均差平方和、积和及自由度，结果如表 7-5 所示。

表 7-5 协方差分析表

变异来源	离均差平方和及积和				估计误差			F
	ν	l_{XX}	l_{XY}	l_{YY}	ν	$\sum (Y - \hat{Y})^2$	MS	
总变异	19	150.55	22.71	4.29	18	0.86		
组间变异	1	17.63	0.89	0.04				
组内变异	18	132.92	21.82	4.25	17	0.67	0.04	
					1	0.19	0.19	4.75

(1) 校正数。

$$C_1 = \frac{213^2}{20} = 2268.45, C_2 = \frac{60.84^2}{20} = 185.08, C_3 = \frac{213 \times 60.84}{20} = 647.95$$

(2) 总的离均差平方和、积和以及自由度。

$$l_{xx} = 2419 - 2268.45 = 150.55, l_{yy} = 189.37 - 185.08 = 4.29$$

$$l_{xy} = 670.66 - 647.95 = 22.71, \nu = 20 - 1 = 19$$

(3) 组间离均差平方和、积和以及自由度。

$$l_{xx} = \left(\frac{76^2}{8} + \frac{137^2}{12} \right) - 2268.4 = 17.68$$

$$l_{yy} = \left(\frac{23.87^2}{8} + \frac{36.97^2}{12} \right) - 185.08 = 0.04$$

$$l_{xy} = \left(\frac{76 \times 23.87}{8} + \frac{137 \times 36.97}{12} \right) - 647.954 = 0.89$$

$$\nu = 2 - 1 = 1$$

(4) 组内离均差平方和、积和以及自由度。

$$l_{xx} = 150.55 - 17.68 = 132.87, l_{yy} = 4.29 - 0.04 = 4.25$$

$$l_{xy} = 22.71 - 0.89 = 21.82, \nu = 19 - 1 = 18$$

(5) 计算总的、组内和修正均数的估计误差平方和、自由度、均方以及 F 值。

$$\text{总的: } \sum (Y - \hat{Y})^2 = 4.29 - \frac{22.71^2}{150.55} = 0.86, \nu = 19 - 1 = 18$$

$$\text{组内: } \sum (Y - \hat{Y})^2 = 4.25 - \frac{21.82^2}{132.87} = 0.67, \nu = 18 - 1 = 17, MS = \frac{0.68}{17} = 0.04$$

$$\text{修正均数: } \sum (Y - \hat{Y})^2 = 0.86 - 0.67 = 0.19, \nu = 18 - 17 = 1, MS = \frac{0.19}{1} = 0.19$$

$$F = \frac{0.19}{0.04} = 4.75$$

(6) $\nu_1 = 1, \nu_2 = 17$ 查 F 界值表, 得 $P < 0.05$ 。说明两组儿童在扣除年龄影响后总体的尿肌酐量均数有差异。

(7) 计算公共回归系数以及各修正均数。

$$b_c = \frac{21.82}{132.92} = 0.16$$

$$\text{正常儿童组: } \bar{Y}_2^* = 2.98 - 0.16 \times (9.50 - 10.46) = 3.13。$$

$$\text{大骨节患儿组: } \bar{Y}_2^* = 3.08 - 0.16 \times (11.42 - 10.46) = 2.93。$$

可以看到在扣除年龄因素之后, 正常儿童总体的尿肌酐含量要高于大骨节病患儿。

分析该资料的程序名为 nkkysj7_1. SAS, 程序如下:

```
data nkkysj7_1;
do c=1 to 2;
input n;
do i=1 to n;
input x y@@ ;
output;
end;
end;
cards;
8
13 3.54 11 3.01 9 3.09 6 2.48 8 2.56 10 3.36 12 3.18 7 2.65
```

```

12
10 3.01 92.83 11 2.92 12 3.09 15 3.98 16 3.89 8 2.21 7 2.39 10 2.74 15 3.36 133.54
11 3.01
;
run;
ods html;
proc glm;
class c;
model y=c x x* c/ssl;
run;
proc glm;
class c;
model y=c x/solution ss3;
lsmeans c/stderr pdiff;
run;
odshtml close;

```

SAS 程序解释：

数据使用循环语句读入。由于两组样本量不相同，用 n 标示不同组中的样本读入的数量，也就是第一组 8 例，第二组 12 例，逐组读入 x 和 y 的值。程序中需要出现 output 语句。

协方差过程采用 proc glm 进行，class 指定分组变量，model 说明模型。上面写了两个过程步，第一个过程步的目的是检验两组间 Y 依赖 X 变化的直线回归方程是否平行。

选择项 ss3 要求系统采用第 III 型离差平方和计算法(与 model 语句中因素的先后顺序无关)；选择项 solution 说明要输出的回归方程中各参数的估计值，Lsmeans(least square means)说明要输出各组 c 的修正均数，选择项 stderr 和 pdiff 指出标准误及均值之间差异比较的概率。

SAS 程序的结果：

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3.66593582	1.22197861	31.12	<0.0001
Error	16	0.62818418	0.03926151		
Corrected Total	19	4.29412000			

R-Square	Coeff Var	Root MSE	y Mean
0.853711	6.513648	0.198145	3.042000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
c	1	0.04524083	0.04524083	1.15	0.2990
x	1	3.58231048	3.58231048	91.24	<0.0001
x*c	1	0.03838451	0.03838451	0.98	0.3375

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3.62755131	1.81377565	46.26	<0.0001
Error	17	0.66656869	0.03920992		
Corrected Total	19	4.29412000			

R-Square	Coeff Var	Root MSE	y Mean
0.844772	6.509367	0.198015	3.042000

这部分结果表明：两条回归直线的斜率是相等的($F = 46.26$, $P < 0.0001$)。它是资料可否进行协方差分析的重要条件之一。

Source	DF	Type III SS	Mean Square	F Value	Pr > F
c	1	0.20061145	0.20061145	5.12	0.0371
x	1	3.58231048	3.58231048	91.36	<0.0001

这部分结果显示： c 和 x 行中的各个数值分别对应的是修正均数的平方和与均方、 F 值以及概率值。由于 P 均小于 0.05，说明两组修正均数间差别有统计学意义。

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	1.206567398	B	0.20424836	5.91	<0.0001
c 1	0.217574451	B	0.09618958	2.26	0.0371
c 2	0.000000000	B	.	.	.
x	0.164169279		0.01717546	9.56	<.0001

这部分结果显示：截距、变量 c 及 x 的参数估计及其检验结果。其中， $x = 0.164169279$ 是公共回归线的斜率。矩阵 $X'X$ 为奇异矩阵，正规方程用广义逆矩阵求解。 B 的估计值是有偏的，不是参数唯一估计值。经 t 检验截距和斜率 P 均 < 0.05 ，有统计学意义。

Least Squares Means					
c	y LSMEAN	Standard Error	H0:LSMEAN=0	H0:LSMean1=LSMean2	
				Pr > t	Pr > t
1	3.17254467	0.07274182	<0.0001	0.0371	
2	2.95497022	0.05865906	<0.0001		

均数总体无效假设概率为 0.0371，有统计学意义。认为在 $\alpha = 0.05$ 水平拒绝 H_0 ，可认为在扣除年龄影响后两种儿童群体的尿肌酐含量不同。程序得到数据和笔算数据的出入是由于小数点后保留位数不同所导致的。

【例 7-2】 研究男儿童体的体表面积是否相同。考虑到儿童的身高和体重对表面积可能有影响，在某地测量了男女各 15 名出生至 3 周岁儿童的身高、体重和体表面积，得到测量数据如表 7-6 所示。

成组设计一元定量资料的协方差分析，以儿童的年龄为协变量 X 做协方差分析。

H_0 ：男儿童体体表面积的总体修正均数相等。

H_1 ：男儿童体体表面积的总体修正均数不等。

$\alpha = 0.05$ 。

表 7-6 3 周岁男女儿童的身高、体重和体表面积

男			女		
身高 (cm)	体重 (kg)	表面积 (cm ²)	身高 (cm)	体重 (kg)	表面积 (cm ²)
54.0	3.00	2446.2	54.0	3.00	2117.3
50.5	2.25	1928.4	53.0	2.25	2200.2
92.0	11.0	5283.3	92.0	12.0	5299.4
94.0	15.0	6101.6	91.0	12.5	5291.5

分析该资料的程序名为 nkkysj7_2.SAS, 程序如下:

```

Data nkkysj7_2;
  do i=1 to 15;
    do sex=1 to 2;
      input x1 x2 y@@ ;
      output;
    end;
  end;
cards;
54.0 3.00 2446.2  54.0 3.00 2117.3
50.5 2.25 1928.4  53.0 2.25 2200.2
51.0 2.50 2094.5  51.5 2.50 1906.2
56.5 3.50 2506.7  51.0 3.00 1850.3
52.0 3.00 2121.0  51.0 3.00 1632.5
76.0 9.50 3845.9  77.0 7.50 3934.0
80.0 9.00 4380.8  77.0 10.0 4180.4
74.0 9.50 4314.2  77.0 9.50 4246.1
80.0 9.00 4078.4  74.0 9.00 3358.8
76.0 8.00 4134.5  73.0 7.50 3809.7
96.0 13.5 5830.2  91.0 12.0 5358.4
97.0 14.0 6013.6  91.0 13.0 5601.7
99.0 16.0 6410.6  94.0 15.0 6074.9
92.0 11.0 5283.3  92.0 12.0 5299.4
94.0 15.0 6101.6  91.0 12.5 5291.5
;
ods html;
proc glm data=nkkysj7_2;
  class sex;
  model y=sex x1 x2/solution;
  lsmeans sex/stderr tdiff;
run;
ods html close;

```

程序说明: 例中, 有两个协变量 x_1 和 x_2 , 一个分组变量 sex 的完全随机化设计的多元协方差分析。data 步为了便于读入数据, 使用了两个 do-end 语句, sex 分组变量令 1 为男, 2 为女。model 语句中, 等号左侧为观察指标, 等号右侧为分组变量和协变量, solution 选项用于求回归方程的系数估计。Lsmeans 语句求修正后均值, stderr 选项求均值的标准误差, tdiff 选项求均值对比的 t 值和 P 值。

程序输出的主要结果:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	68523072.11	22841024.04	557.41	<0.0001
Error	26	1065399.76	40976.91		
Corrected Total	29	69588471.87			

R-Square	Coeff Var	Root MSE	y Mean
0.984690	5.131187	202.4276	3945.043

Source	DF	Type I SS	Mean Square	F Value	Pr > F
sex	1	714100.41	714100.41	17.43	0.0003
x1	1	67440016.92	67440016.92	1645.81	<0.0001
x2	1	368954.79	368954.79	9.00	0.0059

Source	DF	Type III SS	Mean Square	F Value	Pr > F
sex	1	139769.3397	139769.3397	3.41	0.0762
x1	1	938153.7036	938153.7036	22.89	<0.0001
x2	1	368954.7895	368954.7895	9.00	0.0059

由类型 3 的平方和计算结果表明，身高、体重对体表面积影响均有统计学意义。

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	-1255.559200	B	493.5333290	-2.54	0.0172
sex 1	136.828607	B	74.0867551	1.85	0.0762
sex 2	0.000000	B	.	.	.
x1	54.477217		11.3853803	4.78	<0.0001
x2	130.645108		43.5387744	3.00	0.0059

由回归结果可知，x1，x2 相对应的公共偏回归系数分别为 54.477217 和 130.645108，它们与 0 之间的差别的检验结果为 $P < 0.0001$ 和 $P = 0.0059$ 。

Least Squares Means

sex	y LSMEAN	Standard Error	H0:LSMEAN=0	H0:LSMean1=LSMean2	
			Pr > t	t Value	Pr > t
1	4013.45764	52.32694	<0.0001	1.85	0.0762
2	3876.62903	52.32694	<0.0001		

男、女两性体表面积的修正值分别为 52.32694 和 52.32694，两者之间的差别无统计学意义($P = 0.0762$)。

(孔雨佳 王素珍 黄天资)

第 8 章 多重线性回归分析

在内科科研资料的统计分析中,常常需要分析变量之间的相互依赖关系,以说明相依和依存变化的数量关系,回归分析和相关分析是常用的方法。简单直线回归是研究一个因变量与一个自变量呈线性关系的一种统计分析方法。然而,在实际的研究中,事物间的联系是错综复杂的,一个因变量往往受到多个自变量的影响。

将单个自变量的线性关系加以推广,用回归方程定量地刻画一个因变量 Y 与多个自变量 X_1, X_2, \dots, X_p 间的线性依存关系,称为多重线性回归(multiple linear regression)分析。其中,自变量的数值可以自由地随机变动,也可以是人为选定的,但因变量一定是随机的。若所有自变量和因变量都是随机变动的,则既可以做多重回归分析,也可用多重相关(multiple correlation)分析来描述因变量和这一组自变量之间的线性相关性,用偏相关(partial correlation)分析来描述因变量和这一组自变量中的某一个自变量之间,扣除其他变量影响之后的线性相关性。多重线性回归可以用以筛选对因变量有统计学意义的影响因素、估计与预测因变量的取值,还可在给定因变量的取值范围条件下对自变量的取值进行统计控制。

多元线性回归分析在医学上有着广泛的应用,可以研究疾病的致病原因、在临床试验中控制混杂因素,也可以通过建立回归方程进行估计与预测,还可以用于统计控制等。根据不同的需要,选择方程的标准不尽相同。

8.1 多重线性回归方程的建立及假设检验

8.1.1 多重线性回归模型

假定有由 n 个观测构成的随机样本 $(X_{i1}, X_{i2}, \dots, X_{ip}, Y_i), i = 1, 2, \dots, n$, 数据结构如表 8-1 所示。

表 8-1 数据结构

编号	X_1	X_2	...	X_p	Y
1	X_{11}	X_{12}	...	X_{1p}	Y_1
2	X_{21}	X_{22}	...	X_{2p}	Y_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	X_{n1}	X_{n2}	...	X_{np}	Y_n

因变量 Y 和 p 个自变量(也称协变量) X_1, X_2, \dots, X_p 间满足线性回归关系,则多重线性回归模型为:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon$$

(8-1)

其中, β_0 为常数项(constant term), 又称截距(intercept), $\beta_0, \beta_1, \dots, \beta_p$ 称为偏回归系数(partial regression coefficient)或简称回归系数, 表示在其他自变量保持不变时, X_i 增加或减少一个单位时 Y 的平均变化量。 ε 是去除 p 个自变量对 Y 影响后的随机误差, 也称残差(residual), 服从均数为 0、方差为 σ^2 的正态分布。各观测 Y 相互独立, 并且对于每一组 $(X_{i1}, X_{i2}, \dots, X_{ip})$, 因变量 Y 具有相同的方差。

$\beta_0, \beta_1, \dots, \beta_p$ 一般未知, 但可根据样本资料拟合回归方程得到其估计值, 并写出相应的样本回归方程:

$$\hat{Y} = b_0 + b_1X_1 + \dots + b_pX_p \tag{8-2}$$

其中, b_0, b_1, \dots, b_p 是 $\beta_0, \beta_1, \dots, \beta_p$ 的估计值, 也称样本偏回归系数, 简称偏回归系数。

8.1.2 多重线性回归方程的建立及假设检验

建立多重线性回归方程的一般步骤如下。

1) 根据样本数据建立多重线性回归方程

$$\hat{Y} = b_0 + b_1X_1 + \dots + b_pX_p \tag{8-3}$$

其中, $b_0, b_1, b_2, \dots, b_p$ 为模型参数 $\beta_0, \beta_1, \dots, \beta_p$ 的样本估计值, \hat{Y} 为 Y 的估计值。

与简单线性回归分析相同, 多重线性回归模型的参数估计可用最小二乘法得到, 即估计值 \hat{Y} 和实际观察值 Y 的残差平方和 Q 达到最小, 其中

$$Q = \sum_{i=1}^n (Y_i - \hat{Y})^2 = \sum_{i=1}^n [Y_i - (b_0 + b_1X_{i1} + b_2X_{i2} + \dots + b_pX_{ip})]^2$$

令导函数为零, 得到一组线性正规方程组:

$$\begin{cases} b_1l_{11} + b_2l_{12} + \dots + b_pl_{1p} = l_{1Y} \\ b_1l_{21} + b_2l_{22} + \dots + b_pl_{2p} = l_{2Y} \\ \vdots \\ b_1l_{p1} + b_2l_{p2} + \dots + b_pl_{pp} = l_{pY} \end{cases} \tag{8-4}$$

方程组(8-4)中的 l_{ij} 为两个自变量的离均差积和, l_{1Y} 为自变量 X_1 与因变量 Y 的离均差积和。

$$l_{ij} = l_{ji} = \sum (X_i - \bar{X}_i)(X_j - \bar{X}_j) = \sum X_iX_j - \frac{\sum X_i \sum X_j}{n}, i, j = 1, 2, \dots, p \tag{8-5}$$

$$l_{iy} = \sum (X_i - \bar{X}_i)(Y - \bar{Y}) = \sum X_iY - \frac{\sum X_i \sum Y}{n}, j = 1, 2, \dots, p \tag{8-6}$$

解方程组, 可求出回归系数。具体计算复杂, 可借助计算机软件完成。

多重线性回归方程中的常数项估计值:

$$b_0 = \bar{Y} - (b_1\bar{X}_1 + b_2\bar{X}_2 + \dots + b_p\bar{X}_p) \tag{8-7}$$

2) 对多重线性回归方程的假设检验

对多重线性回归方程的假设检验包括对整个模型的假设检验和对偏回归系数的假设检验。

回归模型检验假设如下。

$$H_0: \beta_0 = \beta_1 = \dots = \beta_p = 0。$$

$$H_1: \beta_0, \beta_1, \dots, \beta_p \text{ 不全为零。}$$

因变量 Y 的总变异分解成两部分, 即

$$\sum (Y - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum (Y - \hat{Y})^2 \tag{8-8}$$

其中, $\sum (\hat{Y} - \bar{Y})^2$ 为回归平方和, $\sum (Y - \hat{Y})^2$ 为残差平方和, 上式即

$$SS_{\text{总}} = SS_{\text{回}} + SS_{\text{残}}$$

其中, 回归平方和可用下面公式计算:

$$SS_{\text{回}} = b_1l_{1Y} + b_2l_{2Y} + \dots + b_ml_{mY} = \sum b_jl_{jY} \tag{8-9}$$

残差平方和:

$$SS_{残} = SS_{总} - SS_{回} \tag{8-10}$$

检验统计量:

$$F = \frac{SS_{回}/p}{SS_{残}/(n-p-1)} = \frac{MS_{回}}{MS_{残}} \tag{8-11}$$

统计量 F 服从 $F(p, n-p-1)$ 分布。方差分析见表 8-2。

表 8-2 多重线性回归方差分析表

变异来源	自由度	SS	MS	F	P
总变异	$n-1$	$SS_{总}$			
回 归	p	$SS_{回}$	$SS_{回}/p$	$MS_{回}/MS_{残}$	
残 差	$n-p-1$	$SS_{残}$	$SS_{残}/(n-p-1)$		

如果 $F \geq F_{\alpha, (p, n-p-1)}$, 则在 α 水平上拒绝 H_0 , 接受 H_1 , 认为因变量 Y 与 m 个自变量 X_1, X_2, \dots, X_p 之间存在线性回归关系。

3) 对偏回归系数的假设检验

在医学研究中, 除了要考虑将模型作为一个整体与因变量的相关程度及解释能力外, 往往更关心单个自变量的解释。这就需对其偏回归系数 β_i 进行假设检验。单个回归系数的检验表示的是其他 $p-1$ 个自变量均在当前回归模型中存在的条件下, X_i 的回归系数 β_i 是否为 0 的统计检验, 常用的方法有方差分析和 t 检验。

(1) 方差分析法: 回归方程中某一自变量 X_i 的偏回归平方和 (sum of squares for partial regression), 表示模型中含有其他 $p-1$ 个自变量的条件下该自变量对 Y 的回归贡献, 相当于从回归方程中剔除 X_i 后所引起的回归平方和的减少量, 或在 p 个自变量的基础上新增加 X_i 引起的回归平方和的增加量。检验某一自变量 X_j 与因变量 Y 之间是否具有线性关系, 可用该自变量的偏回归平方和检验 $H_0: \beta_j = 0, H_1: \beta_j \neq 0$ 。

检验统计量:

$$F_i = \frac{SS_{回}(X_i)/1}{SS_{残}/(n-p-1)} \tag{8-12}$$

如果 $F_j \geq F_{\alpha, (1, n-m-1)}$, 则在 α 水平上拒绝 H_0 , 接受 H_1 , 认为因变量 Y 与 X_j 有线性关系。

(2) t 检验法:

$$t_i = \frac{|b_i - 0|}{S_{b_i}} \tag{8-13}$$

b_i 为偏回归系数的估计值, S_{b_i} 为 b_i 的标准误。 t_i 服从自由度 $\nu = n-p-1$ 的 t 分布, 如果 $|t_j| \geq t_{\alpha/2, n-p-1}$, 则在 α 水平上拒绝 H_0 , 接受 H_1 , 认为 Y 与 X_j 有线性回归关系。

8.1.3 多重回归方程的评价

通常用于评价多重线性回归方程拟合效果的统计量有决定系数 (coefficient of determination), 其计算公式为:

$$R^2 = \frac{SS_{回}}{SS_{总}} = 1 - \frac{SS_{残}}{SS_{总}} \tag{8-14}$$

$0 \leq R^2 \leq 1$, 自变量能够解释 Y 变异的百分比, 其值越接近 1, 说明模型对数据的拟合效果越好。

然而, 当方程中包含很多自变量时, 即使其中一些自变量在解释因变量 Y 的变异时贡献很小, 但随着回归方程中自变量的增加, 决定系数仍然会表现为只增不减, 故为了消除自变量个数的影响, 通常计算校正决定系数 (adjusted coefficient of determination), 计算公式如下:

$$R^2_{adj} = \frac{SS_{回}}{SS_{总}} = 1 - \frac{SS_{残}}{SS_{总}} = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1} \tag{8-15}$$

校正决定系数 R^2_{adj} 越大, 说明回归效果越好。当方程中加入有统计学意义的自变量时, R^2_{adj} 增大, 剩余标准差减小; 反之, 若方程中引入无统计学意义的自变量时, R^2_{adj} 可能减小, 而剩余标准差增大。

此外, 复相关系数(multiple correlation coefficient) R 可用来度量因变量 Y 与多个自变量间的线性相关程度, 即观察值 Y 与估计值 \hat{Y} 之间的相关程度。复相关系数计算公式为

$$R = \sqrt{R^2} \tag{8-16}$$

如只有一个自变量时, $R = |r|$, r 为简单相关系数。

8.2 多重回归分析中自变量的筛选

对于多因素线性回归模型而言, 应考虑如何选择自变量进入模型使该模型的近似程度达到最佳。进行变量筛选时, 如果自变量不是特别多, 尽量将所有自变量参与变量筛选, 当自变量特别多且样本量不是很大时, 可以将单个自变量筛选中 $P > 0.05$ 的自变量先不纳入, 必要时随机选择几个参与自变量的筛选。最好选择多种筛选自变量的方法处理同一个多重回归分析资料。

将所有自变量独立与因变量进行简单线性回归分析, 通过假设检验, 挑选出有统计学意义的变量, 并没有进行多重线性回归分析, 直接得出结论; 或者在进行单自变量筛选的基础上, 选择一种方式进行多重回归分析的变量筛选, 以此得出结论都是错误的。因为这会导致一些单独影响较小但当与其他自变量一起时发挥较大作用的自变量的选入, 以及一些在多重回归模型中可能有统计学意义的自变量因某些未知的原因被错误地剔除, 并且后面的方法还可能由于某种筛选变量的方法对于特定资料的极端不吻合而导致错误。

选择自变量的方法有多种, 其基本思路是: 尽可能将回归效果有统计学意义的自变量选入回归方程, 将作用无统计学意义的自变量排除在外。选择自变量的统计学标准主要有全局择优法和逐步选择法。全局择优法是对自变量各种不同的组合所建立的回归方程进行比较, 从全部组合中挑出一个“最优”的回归方程, 如残差平方和最小、残差的均方或剩余标准差最小、 C_p 统计量最小、 U_p 统计量最大等, 虽然效果较好, 但当自变量较多时运算量大。局部最优法可以克服全局择优法当自变量个数较多时计算量大的缺点, 是实际应用中普遍使用的一类方法。

在筛选自变量时, 几种方法各有优劣。全局择优法能得到最优方程, 但计算量大。前进法能使因素的单个作用最大, 但没有考虑交互作用。后退法考虑到变量之间的相互作用, 但当变量间关联强时可能得不到方程。逐步回归综合考虑了因素的单个作用和相互作用, 然而因为某个自变量与因变量相关性较弱可能与其他变量相关, 所以会导致一些与因变量相关的变量未纳入回归方程。实验性研究的统计分析一般不做逐步回归, 并且若使用多重线性回归寻找最佳预测模型, 应该进行残差分析来考察自变量与残差之间是否存在明显的趋势变化。

实际应用中, 直接根据回归诊断和变量筛选之后确定的回归方程参数估计值的大小进行自变量对于因变量作用大小的评价, 由于自变量的不同量纲可能导致估计结果受影响, 应该使用标准化的回归系数的绝对值衡量自变量对于因变量的相对重要性。SAS 中, 在 MODEL 中增加选项 STB 实现。

8.2.1 全局择优法

等价于复相关系数或决定系数最大。一般而言, 模型中的自变量越多, 残差平方和就越小, 即

便引入的是无统计意义的变量, 所以这个准则存在缺陷。对于在相同的自变量个数的情况下, 可用最小平方和准则选择对因变量影响最大的自变量归入模型。

1. 残差的均方或剩余标准差最小

$$MS = \frac{SS_{\text{残}}}{n - p - 1} \quad (8-17)$$

从残差均方的表达式可以看出, 自变量增加可使分子位置减少, 同时分母也减少。 p 很小时, $SS_{\text{残}}$ 和 $n - p - 1$ 均大; p 很大时, $SS_{\text{残}}$ 和 $n - p - 1$ 均小, 欲使残差均方最小, 变量 p 不应过小或过大。等价于校正决定系数 (adjusted coefficient of determination) R_{adj}^2 。有常数项时, R_{adj}^2 定义见式 (8-18a)。定义为:

$$R_{\text{adj}}^2 = 1 - \frac{MS_{\text{残}}}{MS_{\text{总}}} = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} \quad (8-18a)$$

无常数项时, R_{adj}^2 定义见式 (8-18b)

$$R_{\text{adj}}^2 = 1 - \frac{MS_{\text{残}}}{MS_{\text{总}}} = 1 - (1 - R^2) \frac{n}{n - p} \quad (8-18b)$$

R_{adj}^2 开方后得到校正复相关系数 (adjusted multiple correlation coefficient) R_{adj} , 此类统计量评价拟合效果较残差平方和最小要更合理。但当样本量很大时, 使残差均方达到最小的回归模型中, 往往也存在某些自变量没有统计学意义。读者应根据实际情况, 以合适的统计学标准选择自变量进入回归模型。

2. C_p 统计量最小

$$C_p = (n - p - 1) \left(\frac{MS_{\text{残}, p}}{MS_{\text{残}, \text{全}}} - 1 \right) + (p + 1) \quad (8-19)$$

C_p 统计量于 1973 年由 Mallows C. L 提出。式 (8-19) 中, $MS_{\text{残}, p}$ 表示 p 个自变量作回归时的残差均方, $MS_{\text{残}, \text{全}}$ 表示用全部自变量作回归时的残差均方。当入选变量增加时, 式中第一项变小, 第二项变大。想得到较小的 C_p 值, 必须是这两项的折中: 一方面, 入选的自变量不能太多, 以使方程易于解释并降低工作量及研究费用; 另一方面, 自变量也不能太少, 以保证一定的估计和预测精度。因此, C_p 统计量最小的标准相对得到较多推崇。一般认为: 当 C_p 的值等于或近似等于回归方程中自变量的个数时, 表明此时的回归方程对资料的拟合效果最好。

3. U_p 统计量最大

$$U_p = \frac{1 - R^2}{(n - p - 1)(n - p - 2)} \quad (8-20)$$

入选自变量 p 很小时, 分子分母均大; 入选自变量 p 很大时, 分子分母均小。欲使 U_p 最大, 变量个数 p 必须适中。

4. AIC 统计量最小

$$AIC = n \ln \left(\frac{SS_{\text{差}}}{n} \right) + 2m \quad (8-21)$$

8.2.2 局部选择法

用全局择优法建立的最优回归方程在估计与预测中的效果最好, 然而当自变量数量较大时, 计算量就变得非常大。逐步选择法可以克服采用全局择优法当自变量个数较多时计算量大的缺点, 是实际应用中普遍使用的一类方法。按照入选变量的顺序不同分为前进法 (forward selection)、后退

法(backward elimination)和逐步回归法(stepwise regression)。它们的共同特点是依次分析所有可能的模型,每一步只纳入或剔除一个自变量,逐步达到最优模型的条件,其中决定取舍的标准是基于对偏回归平方和的 F 检验 ($H_0: \beta_j = 0, H_1: \beta_j \neq 0$)

$$F = \frac{SS_{\text{回}}^{(i)}(X_j)}{SS_{\text{残}}^{(i)} / (n - p - 1)} \tag{8-22}$$

1. 前进法 (forward)

前进法是在建立线性方程时从无到有、从少到多地纳入自变量。首先,应对所有的自变量做线性回归,对偏回归平方和最大的自变量进行 F 检验,若其偏回归系数有统计学意义,也就是 P 值小于 $SLENTRY$ 时将自变量纳入方程。然后,在已有一个自变量的基础上,对剩余的偏回归平方和最大的所有自变量进行 F 检验以决定是否再次纳入方程。重复这个过程,直到剩余的变量中没有可以再纳入方程的自变量为止。前进法可以自动纳入高度相关的自变量,却会在纳入新变量时可能会使原先已经纳入方程的变量变得不重要。

2. 后退法 (backward)

后退法与前进法相反,是从包含全部变量的模型开始分析,逐步剔除对方程无统计学意义的变量。剔除的时候,首先选择纳入方程的所有变量中偏回归平方和最小的变量通过 F 检验进行分析,若无统计学意义,则剔除。然后,对方程的剩余因变量进行分析,每次都选择偏回归平方和最小的进行 F 检验逐一剔除值小于 $SLSTAY$ 的变量。重复这个过程,当方程中的所有因变量都不能够剔除为止。后退法考虑到了自变量的组合作用,然而当变量之间高度相关时可能出现错误的结果。

3. 逐步回归法 (stepwise)

逐步回归法是在上述两种方法的基础上,进行双向筛选的一种方法。这种方法本质上是前进法,也是将所有变量从无到有、从少到多地纳入方程,只是在所有未纳入方程的变量中选择偏回归平方和最大的变量进行 F 检验,按 $SLENTRY$ 水平有统计学意义而进入模型的变量后,再在此基础上分析已有所有纳入的所有自变量是否由于新纳入方程的变量而变得“不重要”,也就是选择已经纳入方程的偏回归平方和最小的自变量根据 F 统计量按 $SLSTAY$ 水平决定是否需要进行剔除。不断重复这个过程,直到没有新变量可以纳入方程,也没有变量可以从方程中剔除,得到的方程就是局部最优的回归方程。

α 越小说明选取自变量的标准越严格,能够进入方程的自变量越少,反之,越大的 α 则意味着更多的因素可以进入方程。一般而言,小样本的 α 值可以取 0.10 和 0.15,而大样本的 α 一般定为 0.05。对于纳入和剔除自变量的 F 检验的检验水准,可以设置相同,也可以设置不同。一般而言,选入自变量的检验水准要小于或者等于剔除自变量的检验水准,也就是 $\alpha_{\text{se}} \leq \alpha_{\text{sls}}$ 。

8.3 多重回归分析应用的注意事项

8.3.1 模型的选择

当因变量是非时间连续变量时,选择多重线性回归分析;当因变量是二值变量时,选择多重 logistic 回归分析;当因变量是多值有序变量时,选择有序变量的多重 logistic 回归分析;当因变量是生存时间变量时,选择 Cox 回归分析或参数回归分析;当因变量是时间协变量时,选择时间序列分析。

如果建立的回归模型用于预测,则在建立的回归方程中应该选择 R^2 大的。如果建立的回归方

程式进行逆估计，如控制自变量来保证因变量在一个确定的范围内波动，则除了要求回归方程的 R^2 大外，还需要回归系数的标准误小。

8.3.2 样本含量

多重线性回归方程既可用于大样本资料，又可用于小样本资料。但当样本含量 n 相对于自变量 p 不是很多时，容易造成假象，所以应注意 n 和 p 的比例。有学者认为 n 至少应该为方程中自变量个数的 5 ~ 10 倍，也有学者认为应该使用经验公式 $n \geq 8(1 - R_c^2)/R_c^2 + (p - 1)$ 估计样本含量，其中 R_c^2 为多重回归预期的决定系数。

8.3.3 指标的数量化

对于多重线性回归的资料，一般要求因变量 Y 为连续变量，自变量 X 可以是连续变量、分类变量或有序变量。

(1) 自变量是连续变量，此时连续变量以原始观测值的形式出现。当自变量 X 和因变量 Y 不呈线性关系时，可以考虑对 X 做某种变换，以改善回归方程的拟合优度。

(2) 自变量是分类变量，此时必须对指标进行数量化处理。

当指标是二分类时，可以赋值：

$$X = \begin{cases} 1 & \text{男性} \\ 0 & \text{女性} \end{cases} \quad X = \begin{cases} 0 & \text{男性} \\ 1 & \text{女性} \end{cases}$$

当指标是多分类时，可以通过定义一组变量(哑变量)进行数量化：

$$X_1 = \begin{cases} 1 & \text{A 型} \\ 0 & \text{其他血型} \end{cases} \quad X_2 = \begin{cases} 1 & \text{B 型} \\ 0 & \text{其他血型} \end{cases} \quad X_3 = \begin{cases} 1 & \text{AB 型} \\ 0 & \text{其他血型} \end{cases}$$

此时，有如表 8-3 所示的关系。

表 8-3 多分类名义变量与哑变量的对应关系

哑变量	A 型	B 型	AB 型	O 型
X_1	1	0	0	0
X_2	0	1	0	0
X_3	0	0	1	0

采用不同的赋值方式，得到的回归方程不同，但回归方程的总体效果不改变。需要注意的是，当 $g > 2$ 时，不能用常规的逐步回归方法进行变量选择，而应将 $g - 1$ 个哑变量作为一个整体来考虑是否纳入方程。

(3) 自变量是有序变量，此时有两种赋值方式：一种是根据程度赋分，按连续变量处理；另一种是用哑变量表示。若方程中变量不是很多，并且样本含量又大，用哑变量方法较好；当自变量较多、样本含量不够大，按连续变量处理更合适。

8.3.4 多重线性回归分析中的多重共线性问题和交互作用

多重共线性指的是一些自变量之间存在较强的线性关系，这种情况在实际应用中非常普遍，如研究高血压与年龄、吸烟年限、饮白酒年限等因素的关系。如果这种共线性的相关程度非常高，则可能使得最小二乘法建立的回归方程失效，如参数估计值的标准误变大，回归方程不稳定， t 值不准确，估计值的符号和客观实际不一致等。

对自变量之间是否存在较强的共线性关系进行研究，称为共线性诊断问题(diagnosis of collinearity)。当出现下面情况之一时，则认为存在多重共线性：若两个自变量的相关系数接近 1；若

决定系数 R^2 很大(一般大于 0.8), 但模型中全部或部分偏回归系数的检验并不显著; 若 m 个自变量分别以其中一个对其他所有自变量建立回归方程, 对应 m 个方程的决定系数中最大的接近 1, 则这个自变量与其他自变量的一个或者几个存在强线性相关性; 方差膨胀因子又称为方差膨胀系数或方差膨胀值 (Variance Inflation), $VIF = 1/TOL$, TOL 给出容许度, $TOL = 1 - R^2$, 对于一个人选变量而言, 容许度小表明该自变量与其他自变量之间的关系密切。不好的试验设计方差膨胀因子 $VIF = 1/TOL$ 的取值可能趋于无限大。经验上, $VIT > 5$ 或 10, 可以认为存在严重的共线性。

在 SAS regression 过程的 MODEL 中增加选项 COLLIN(对截距项未校正的共线性诊断)和 COL-LINOINT(对截距项校正的共线性诊断)来实现回归诊断。当截距项无统计学意义时, 依据 COLLIN 结果判断共线性情况, 当截距项有统计学意义时, 应根据 COLLINOINT 结果判断共线性情况。两者都给出了信息矩阵的特征根和条件数, 还有方差分量, 经验上, $0 \leq k < 10$, 则认为没有多重共线性; $10 \leq k < 30$, 则认为中等程序或较强的多重共线性; $k \geq 30$, 则认为自变量之间存在一定程度的相关性。强的多重共线性同时表现在变量的方差分量上, 最大条件数同时有 2 个以上自变量的方差分量超过 50%, 意味着自变量间有一定程度的相关性。方差膨胀因子可利用 SAS regression 过程中 MODEL 的选项 VIF 给出方差膨胀因子。

如果出现多重共线性, 则可以直接删减造成共线性但对应变量影响不大的自变量; 如果使用逐步回归建立回归方程则可以避免自变量之间存在共线性; 另外, 使用互相不相关的少数几个主成分来建立回归方程的主成分回归法也能避免共线性。

对于因变量影响特别大的点称为强影响点或异常点。异常点可能导致结果的偏离甚至错误, 严重影响回归方程的质量, 根据不同情况对诊断出来的异常点进行处理, 可以使用 Cook's D 距离统计量、学生化残差 STUDENT 统计量、RSTUDENT 统计量等, 其中学生化残差统计量使用最为便捷。

当一个自变量对因变量的大小与另一个自变量取值有关时, 表明这两个自变量之间存在交互作用。是否存在交互作用通常依靠专业知识考虑, 而对回归方程添加乘积项考察交互项系数是否为零来判断。

8.3.5 残差分析

如果实际资料与模型假设有较大偏离, 最小二乘估计、假设检验以及区间估计可能出现错误。为了正确使用多重线性回归, 建立模型之前需要检验资料是否满足回归条件, 残差分析 (residual analysis) 是一种检查资料是否满足模型条件的方法。所谓残差, 就是观察值 Y_i 与估计值 \hat{Y}_i 的差值, 即 $e_i = Y_i - \hat{Y}_i$, 且 $e_i \sim N(0, \sigma^2)$ 。

残差图既可以用于考察模型条件是否满足, 还可以深入考察假设的合理性, 也可以用于协助寻找离群点。

使用残差考察模型条件是否成立的方法是: 计算标准化残差 $e'_i = \frac{e_i}{\sqrt{MS_{\text{回}}}}$ (其中, $e'_i \sim N(0, 1)$),

以标准化残差 e'_i 为纵轴, 以 \hat{Y}_i 为横轴绘制残差图。在残差图中, 若 e'_i 点较为均匀的散布在以 0 为中心的兩直线之间, 可以认为对任意的自变量, 因变量有相同的方差; 若 e'_i 随着 \hat{Y}_i 的增大而出现收敛或发散的趋势, 说明方差不一致; 若 e'_i 随着 \hat{Y}_i 表现出线性趋势, 则可能因变量 Y 与自变量之间不是线性关系或者残差 e'_i 不满足独立性。

如果以标准化残差 e'_i 为纵轴, 以某一自变量为横轴建立残差图可以进一步考察相关假设的合理性。另外, 残差图中 $e'_i > 2$ 的点通常看成离群点, 要分析造成离群点原因, 再选择合适的方法进行处理。

8.4 案例分析

【例 8-1】 35 名怀疑患有动脉硬化的就诊患者的载脂蛋白 AI、载脂蛋白 B、载脂蛋白 E、载脂蛋白 C、低密度脂蛋白中的胆固醇、高密度蛋白中的胆固醇含量，如表 8-4 所示。

表 8-4 35 名就诊患者血清中低、高密度脂蛋白中的胆固醇含量及载脂蛋白的测量结果

序号	X ₁ 载脂蛋白 AI(mg/)	X ₂ 载脂蛋白 B(mg/)	X ₃ 载脂蛋白 E(mg/)	X ₄ 载脂蛋白 C(mg/)	Y ₁ 低密度脂蛋白 (mg/)	Y ₂ 高密度脂蛋白 (mg/)
1	173	106	7.0	14.7	137	62
2	139	132	6.4	17.8	162	43
.....						
34	188	120	6.2	17.6	161	59
35	177	109	4.6	28.1	147	75

以 Y₁为因变量建立回归方程，分析该资料的程序名为 nkkysj8_1. SAS。

```
data nkkysj8_1;          * 建立进行多重线性回归分析的数据集;
input x1 x2 x3 x4 y1 y2;
cards;
173 106 7.0 14.7 137 62
139 132 6.4 17.8 162 43
198 112 6.9 16.7 134 81
118 138 7.1 15.7 188 39
139 94 8.6 13.6 138 51
175 160 12.1 20.3 215 65
131 154 11.2 21.5 171 40
158 141 9.7 29.6 148 42
158 137 7.4 18.2 197 56
132 151 7.5 17.2 113 37
162 110 6.0 15.9 145 70
144 113 10.1 42.8 81 41
162 137 7.2 20.7 185 56
169 129 8.5 16.7 157 58
129 138 6.3 10.1 197 47
166 148 11.5 33.4 156 49
185 118 6.0 17.5 156 69
155 121 6.1 20.4 154 57
175 111 4.1 27.2 144 74
136 110 9.4 26.0 90 39
153 133 8.5 16.9 215 65
110 149 9.5 24.7 184 40
160 86 5.3 10.8 118 57
112 123 8.0 16.6 127 34
147 110 8.5 18.4 137 54
204 122 6.1 21.0 126 72
131 102 6.6 13.4 130 51
170 127 8.4 24.7 135 62
173 123 8.7 19.0 188 85
132 131 13.8 29.2 122 38
146 90 8.5 26.8 188 58
```

```
137 112 9.5 25.4 87 37
155 84 5.1 10.9 116 55
188 120 6.2 17.6 161 59
177 109 4.6 28.1 147 75
;
run;
ods html;
proc reg;                                * 使用逐步回归法建立多重线性回归方程;
model y1 =x1 x2 x3 x4/vif selection=stepwise slentry=0.1 slstay=0.05 stb;
model y2 =x1 x2 x3 x4/ vif selection=stepwise slentry=0.1 slstay=0.05 stb;
run;
proc reg;                                * 使用前进法建立多重线性回归方程;
model y1 =x1 x2 x3 x4/ vif selection=forward stb; * vif 表示变量间相关性的方差膨胀系数;
model y2 =x1 x2 x3 x4/ vif selection=forward stb; * stb 表示要求输出标准回归系数;
run;
proc reg;                                * 使用后退法建立多重线性回归方程;
model y1 =x1 x2 x3 x4/ vif selection=backward stb;
model y2 =x1 x2 x3 x4/ vif selection=backward stb;
run;
proc reg;                                * 使用修正 R2 选择法建立多重线性回归方程;
model y1 =x1 x2 x3 x4/vif selection=adjrsq;
model y2 =x1 x2 x3 x4/vif selection=adjrsq;
run;
ods html close;
quit;
```

SAS 运行结果如下。

对 Y_1 建立回归方程。

(1) 逐步回归方法。

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x2		1	0.2346	0.2346	5.9875	10.12	0.0032
2	x4		2	0.1227	0.3573	2.0593	6.11	0.0190

模型的拟合情况。

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	14160	7080.04579	8.90	0.0008
Error	32	25469	795.90250		
Corrected Total	34	39629			

Root MSE	28.21174	R-Square	0.3573
Dependent Mean	149.97143	Adj R-Sq	0.3171
Coeff Var	18.81141		

方程中的参数估计情况，包括截距项和自变量回归系数等的估计量以及对应的假设检验结果。其中，Parameter Estimate 为入选自变量的回归系数的估计，Standard Error 为相应参数的标准误，t value 和 $Pr > |t|$ 分别是针对相应参数为 0 的假设检验结果，Standardized Estimate 为最终选入方程自变量的标准回归系数的估计。

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	66.50751	32.25107	2.06	0.0474	0
x2	1	0.97801	0.25561	3.83	0.0006	0.55233
x4	1	-1.75903	0.71176	-2.47	0.0190	-0.35677

(2)用前进法建立回归方程。

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	28.15338	49.20127	0.57	0.5713	0
x1	1	0.21726	0.21064	1.03	0.3103	0.14816
x2	1	1.02311	0.25907	3.95	0.0004	0.57780
x4	1	-1.79191	0.71176	-2.52	0.0172	-0.36343

(3)用后退法建立回归方程。

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	66.50751	32.25107	2.06	0.0474	0
x2	1	0.97801	0.25561	3.83	0.0006	0.55233
x4	1	-1.75903	0.71176	-2.47	0.0190	-0.35677

(4)修正 R^2 选择法。

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	1	28.15338	49.20127	0.57	0.5713	0
x1	1	0.21726	0.21064	1.03	0.3103	1.02947
x2	1	1.02311	0.25907	3.95	0.0004	1.06803
x4	1	-1.79191	0.71176	-2.52	0.0172	1.03970

以逐步回归法建立 Y_1 多重线性方程为例说明具体的假设检验的步骤如下。

(1)对回归方程做假设检验。

$H_0:\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ H_1 :各 β_i 不全 0 $\alpha = 0.05$

$F = \frac{MS_{回}}{MS_{剩}} = \frac{SS_{回} / \nu_{回}}{SS_{剩} / \nu_{剩}} = 8.90$

得到 $P = 0.0008$ ，按 $\alpha = 0.05$ 的水准，拒绝 H_0 ，各自变量的偏回归系数不全等于 0，多重回归方程成立。

(2)对偏回归系数做假设检验。

判断总的回归关系成立以后，表 8-5 给出了各回归系数的最小二乘法估计。

表 8-5 数据的偏回归系数估计及其假设检验(Y_1 为因变量)

变量	DF	回归系数	标准误	t 值	P 值	标准回归系数
常数项	1	66.50751	32.25107	2.06	0.0474	0
X_2	1	0.97801	0.25561	3.83	0.0006	0.55233
X_4	1	-1.75903	0.71176	-2.47	0.0190	-0.35677

回归方程 $\hat{Y}_1 = 66.50751 + 0.97801X_2 - 1.75903X_4$

标准回归方程 $\hat{Y}_1 = 0.55233X_2 - 0.35677X_4$

可以看到在 $\alpha = 0.05$ 的水平上， X_2 和 X_4 对 Y_1 的回归方程建立具有统计学意义。回归方程表明，载脂蛋白 X_2 和载脂蛋白 X_4 影响低密度脂蛋白 Y_1 ，低密度脂蛋白 Y_1 随着载脂蛋白 X_2 的增加而增大，随着载脂蛋白 X_4 的增加而减小。当载脂蛋白 X_4 固定时，载脂蛋白 X_2 每增加一个单位， Y_1 平均增加了 0.55233 个单位；当载脂蛋白 X_2 固定时，载脂蛋白 X_4 每增加一个单位， Y_1 平均减少了 0.35677 个单位。

根据自变量不同的筛选方法，得到以下回归方程。

逐步回归法 $\hat{Y}_1 = 66.50751 + 0.97801X_2 - 1.75903X_4$

前进法 $\hat{Y}_1 = 28.15338 + 0.21726X_1 + 1.02311X_2 - 1.79191X_4$

后退法 $\hat{Y}_1 = 66.50751 + 0.97801X_2 - 1.75903X_4$

修正 R^2 选择法 $\hat{Y}_1 = 28.15338 + 0.21726X_1 + 1.02311X_2 - 1.79191X_4$

结合全局选择法和局部选择法，最终选定回归方程。

$\hat{Y}_1 = 28.15338 + 0.21726X_1 + 1.02311X_2 - 1.79191X_4$

对 Y_2 建立回归方程。

(1) 逐步回归法。

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	-20.43250	8.91099	-2.29	0.0284	0
x1	1	0.48771	0.05714	8.54	<.0001	0.82961

(2) 前进法。

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	-5.20800	11.42075	-0.46	0.6516	0
x1	1	0.46359	0.05851	7.92	<.0001	0.78859
x3	1	-0.84857	0.71520	-1.19	0.2444	-0.13489
x4	1	-0.23367	0.21331	-1.10	0.2818	-0.11821

(3) 后退法。

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate
Intercept	1	-4.87882	11.45241	-0.43	0.6730	0
x1	1	0.45091	0.05754	7.84	<.0001	0.76701
x3	1	-1.25072	0.61570	-2.03	0.0506	-0.19882

(4) 修正 R^2 选择法。

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	1	-5.20800	11.42075	-0.46	0.6516	0
x1	1	0.46359	0.05851	7.92	<.0001	1.15529
x3	1	-0.84857	0.71520	-1.19	0.2444	1.50718
x4	1	-0.23367	0.21331	-1.10	0.2818	1.35798

用三种方法得到的 Y_2 的回归方程如下。

逐步回归法 $\hat{Y}_2 = -20.43250 + 0.48771X_1$
前进法 $\hat{Y}_2 = -5.20800 + 0.46359X_1 - 0.84857X_3 - 0.23367X_4$
后退法 $\hat{Y}_2 = -4.87882 + 0.45091X_1 - 1.25072X_3$
修正 R^2 选择法 $\hat{Y}_2 = -5.20800 + 0.46359X_1 - 0.84857X_3 - 0.23367X_4$
结合全局选择法和局部选择法, 选择回归方程。

$$\hat{Y}_2 = -5.20800 + 0.46359X_1 - 0.84857X_3 - 0.23367X_4$$

【例 8-2】 原文题目：炎症及免疫指标对类风湿关节炎患者血脂的影响分析。方法：选取 RA 患者 225 例, 检测入院第 1 天的血脂、血沉 (ESR)、C 反应蛋白 (CRP)、类风湿因子 (RF)、抗环瓜氨酸肽抗体 (抗 CCP)、抗角蛋白抗体 (AKA)、抗核周因子抗体 (APF)、补体 (C)。将 logTC、logHDL-C、logLDL-C、logTG 及 AIP 作为因变量, 将上述炎症与免疫指标中与因变量单因素回归分析有统计学意义 ($P < 0.2$) 的指标作为自变量进行多因素逐步回归分析。结果：多因素逐步回归分析显示, C 与 TC 水平正相关 ($R^2 = 0.067, P < 0.05$), ESR、CRP 与 HDL-C 水平负相关 ($R^2 = 0.202, P < 0.05$), CRP、抗 CCP 抗体与 LDL-C 正相关 ($R^2 = 0.129, P < 0.05$), ESR、C 与 AIP 正相关 ($R^2 = 0.046, P < 0.05$)。

【实例辨析】 原文作者在分析血脂、血沉 (ESR)、C 反应蛋白 (CRP)、类风湿因子 (RF)、抗环瓜氨酸肽抗体 (抗 CCP)、抗角蛋白抗体 (AKA)、抗核周因子抗体 (APF) 和补体 (C) 共 8 种炎症及免疫指标对类风湿关节炎患者血脂的影响分析时, 首先将上述炎症与免疫指标与因变量进行单因素回归分析, 然后将有统计学意义 ($P < 0.2$) 的指标再作为自变量进行多因素逐步回归分析, 这样做是不对的。一般情况下, 当自变量不是很多时, 应尽可能地使全部自变量都有机会参与变量的筛选; 当自变量特别多且样本含量较小时 (一般要求样本含量为自变量个数的 5 ~ 10 倍以上, 回归分析的结果才较稳定), 可以将单个自变量筛选中“ $P > 0.5$ ”的那些自变量暂时不参与自变量的筛选过程, 必要时可从其中随机抽取几个与拟参与自变量筛选的那些自变量一并考察。本例自变量有 8 个, 样本含量为 225, 因此在进行多重回归分析时, 应将所有的自变量代入回归方程进行分析。在进行自变量筛选时, 通常采用逐步回归分析法, 但必要时也可多用几种筛选变量的方法, 综合分析, 以便得出较为可靠的结果。

(孔雨佳 赵晓蒙 王素珍)

第9章 Poisson 回归模型

医学科学研究中,特别是流行病学研究中,常需要探索危险因素与发病率较低的疾病(如各类癌症、遗传缺陷)间的关系。此时,数据的统计处理不宜采用常规的 logistic 回归或 Cox 回归等模型。这类疾病的发病例数通常服从 Poisson 分布,因此 Poisson 回归模型是分析流行病学中低发病率队列资料的首选方法。

9.1 Poisson 回归模型及参数估计

Poisson 回归模型是描述服从 Poisson 分布的目标变量 y_i 的均数 μ_i 与协变量 $X_i(x_1, x_2, \dots, x_m)$ 间依赖关系的回归模型。目标变量 y_i 的概率密度函数表示为:

$$f(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, y_i = 1, 2, 3, \dots$$

回归模型可简单记为: $g(\mu) = \beta_0 + X\beta$ 。其中 g 为连接函数。对符合 Poisson 分布的因变量,连接函数一般取为自然对数函数。因此,发病率模型可写为:

$$\log(\mu_i) = \log(n_i) + \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im} \quad (9-1)$$

其中, n_i 为该协变量模式下的总观察人数。

这里,回归模型的似然函数可表示为:

$$L(\beta) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \left(\frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \right) \quad (9-2)$$

通常采用极大似然估计法对参数进行估计。

9.1.1 模型拟合

Poisson 回归模型拟合优劣采用偏差统计量(deviance)表示,它在 Poisson 分布条件下的计算公式为:

$$G^2 = 2 \sum (y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i)) \quad (9-3)$$

G^2 服从 $n-p$ 的 χ^2 分布(p 为模型中的参数个数)。

9.1.2 参数检验

参数检验是比较两个包含不同参数个数的模型的偏差统计量 G^2 (ΔG^2),当 $\Delta G^2 > \chi_{\Delta v}^2$ 时(Δv 表示两个模型自由度之差), $P < 0.05$,该参数有统计学意义。

9.2 实例分析

【例 9-1】2007 年,Edgren 等人对受血者曾输亚临床癌症献血者的血液癌症发病率进行回顾性队列研究。文中定义献血者献血后 5 年内确诊为癌症诊断为亚临床癌症,结果发表于《The Lancet》杂

志上。本研究收集 1968—2002 年瑞典和丹麦血库计算机登记的资料，并进行了资料汇总，部分结果见表 9-1，为校正距离第一次输血时间这个混淆因素，以距离第一次输血时间进行分层。本例中发病率低，试采用 Poisson 回归模型分析年龄效应和暴露组与未暴露组间的差别。

表 9-1 曾输注亚临床癌症献血者的血液的受血者与输注正常献血者的血液的受血者的资料

距离第一次输血 时间(年)	暴露组		未暴露组	
	癌症发病例数	观察人数	癌症发病例数	观察人数
0 ~ 4	511	45181	13412	1313197
5 ~ 9	297	26585	7705	877132
10 ~ 19	148	16387	5874	733825
20 ~ 34	22	2775	1682	185718

表 9-1 中，距离第一次输血时间为分层因素，分为 4 层，是否输注亚临床癌症献血者的血液为暴露因素，分析目的是估计暴露组相对于未暴露组的危险性，以及各年龄组的相对危险性与变化趋势。Poisson 回归模型分析的 SAS 程序名为：nkkysj9_1. SAS。SAS 程序如下：

```
data nkkysj9_1;
  input time treat y n @@ ;
  ln=log(n);
cards;
1 1 511 45181 1 2 13412 1313197 2 1 297 26585 2 2 7705 877132
3 1 148 16387 3 2 5874 733825 4 1 22 2775 4 2 1682 185718
;
run;
proc genmod data = nkkysj9_1;
  class time treat;
  model y=time treat/dist =poisson link=log offset =ln noint;
run;
```

程序说明：本例距离第一次输血时间以 time = 1, 2, 3, 4 表示，是否暴露以 treat = 1, 2 表示，暴露组设为 1，未暴露组设为 2。数据集中第一列为距离第一次输血时间，第二列为处理(是否暴露)，第三、四列为发病人数与观察人数。调用 SAS 中 genmod 过程进行 Poisson 回归模型分析，具体参数设置参考 SAS 帮助。

输出主要结果：
(第一部分)

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	3	5.3410	1.7803
Scaled Deviance	3	5.3410	1.7803
Pearson Chi – Square	3	5.2853	1.7618
Scaled Pearson X2	3	5.2853	1.7618
Log Likelihood		235924.7159	
Algorithm converged.			

(第二部分)

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	0	0.0000	0.0000	0.0000 0.0000	.	.

time	1	1	-4.5854	0.0086	-4.6022	-4.5687	286641	<.0001
time	2	1	-4.7312	0.0112	-4.7532	-4.7092	177417	<.0001
time	3	1	-4.8282	0.0129	-4.8535	-4.8029	139823	<.0001
time	4	1	-4.7083	0.0242	-4.7558	-4.6608	37754.6	<.0001
treat	1	1	0.1387	0.0325	0.0749	0.2025	18.17	<.0001
treat	2	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

结果解释:

第一部分显示模型拟合情况，结果显示计算收敛，本例中模型拟合指标中偏差统计量 G^2 等于 5.3410,(自由度 $\nu = 3$)， $P > 0.05$ ，这说明模型拟合良好。第二部分为参数估计，结果包括参数及参数标准误估计值、95% 置信区间、参数检验卡方值及对应的 P 值。时间和是否暴露均有统计学意义，暴露组相对于未暴露组的发病危险度为 $\exp(0.1387) = 1.149$ 。不同年龄组参数估计值结果显示，随着时间增加，发病危险性逐渐减小。

(曹文君 黄天资)

第 10 章 logistic 回归

多重 logistic 回归分析和多重线性回归分析都可以分析一个因变量与多个自变量之间的依存关系，可以解决筛选自变量和更精确地对因变量做预测等问题，但线性回归要求反应变量或因变量是连续型变量，而 logistic 回归分析可以处理医学研究实践中因变量为二项分类、多项有序分类和多项无序分类变量时研究反应变量和诸多自变量间的相互关系问题。多重 logistic 回归分析属于概率型非线性回归模型，可以应用于流行病学中危险因素的筛选、病因学分析或预测发病概率；分析药物或毒物的剂量效应关系。在分析疾病与危险因素定量关系的流行病学研究中，需要排除混杂因素的影响，当样本量大且分析因素少时可以使用 Mantel-Haenszel 分层分析法。然而，若此条件不满足，则可选择 logistic 回归来排除混杂因素的影响。线性回归中，当因变量为 0,1 取值，各因素都处于低水平或高水平时，因变量的预测值可能超出 0 和 1 的范围而出现不合理现象时也可以使用 logistic 回归来解决。

10.1 logistic 回归

10.1.1 二分类反应变量的 logistic 回归分析

logistic 回归模型 设因变量 Y 是一个二值变量，取值为：

$$Y = \begin{cases} 1 & \text{出现阳性结果} \\ 0 & \text{出现阴性结果} \end{cases}$$

影响反应变量 Y 的因素有 p 个，称为自变量 (independent variable) 或解释变量 (explanatory variable)，记为 X_1, X_2, \dots, X_p 。

logistic 回归模型：

$$P = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)} \quad (10-1)$$

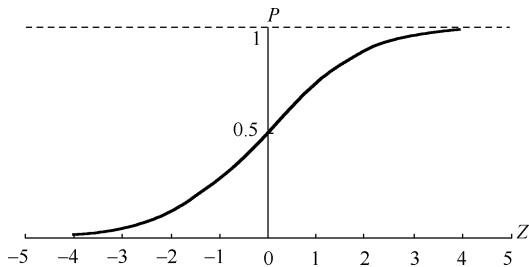
其中， β_0 为常数项或截距， $\beta_1, \beta_2, \dots, \beta_p$ 称为 logistic 回归模型的回归系数。logistic 模型是一个概率型非线性的回归模型，自变量可任意取值，既可以是数值变量，也可以是分类变量经赋值后的哑变量 (dummy variable)。当自变量为多水平的名义变量时，需要转换成 (水平数 - 1) 个哑变量来参与分析。当自变量为连续型变量时，是否转化成分组变量以及何时转化需要根据实际研究目的来确定。

令 $Z = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$ ， Z 趋于 $+\infty$ ， P 值趋近于 1； Z 近于 $-\infty$ ， P 值趋近于 0。 P 和 Z 的关系如图 10-1 所示， P 值在 $[0, 1]$ 间以点 $(0, 0.5)$ 为中心呈对称 S 形变化。

对公式 (10-1) 进行对数变换，又称 logit 变换，可以得到

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (10-2)$$

转化之后，等式左侧的 $\text{logit}(P)$ 是阳性结果发生概率与阴性结果发生概率之比的自然对数，其取值范围为 $(-\infty, +\infty)$ ，也就是说 $(-\infty, +\infty)$ 间任何 $\ln[P/(1-P)]$ 值，均能与 $(0, 1)$ 之间的 P 值对应。

图 10-1 logistic 回归中 P 和 Z 的关系图

根据样本数据可以估计出 logistic 回归模型的常数项 b_0 和回归系数 (b_1, b_2, \dots, b_p) , 从而描述和分析反应变量与自变量间的关系, 并且可以计算在特定条件下阳性结果发生的概率。

常见的 logistic 回归方程还有以下几种等价形式:

$$\frac{P}{1-P} = e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p} \quad (10-3)$$

$$P = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}} \quad (10-4)$$

$$1 - P = \frac{1}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}} \quad (10-5)$$

10.1.2 logistic 回归模型的参数估计和假设检验

1. 参数估计与优势比估计

1) 参数估计

logistic 模型的建立需要根据样本数据计算出参数 β_0 和 β_i , 通常用最大似然估计(maximum likelihood estimate, MLE), 具体过程如下。

首先, 建立样本似然函数:

$$L = \prod_{i=1}^n P_i^{Y_i} (1 - P_i)^{1-Y_i} \quad (10-6)$$

式中, P_i 表示第 i 例观察对象在暴露条件下阳性结果发生的概率, 如果实际出现的是阳性结果, 取 $Y_i = 1$, 否则取 $Y_i = 0$ 。根据极大似然估计的原理, 在一次抽样中获取现有样本概率应该最大, 也就是样本似然函数 L 应该取得最大值。

为了便于计算, 取似然函数的对数形式:

$$\ln L = \sum_{i=1}^n [Y_i \ln P_i + (1 - Y_i) \ln(1 - P_i)] \quad (10-7)$$

用 Newton-Raphson 迭代方法求使对数似然函数达到极大值时 $\beta_0, \beta_1, \dots, \beta_p$ 的最大似然估计值 b_0, b_1, \dots, b_p , 同时得到参数估计值的方差—协方差矩阵(对角线开平方为标准误 $S_{b_0}, S_{b_1}, \dots, S_{b_p}$)。

如果是分层资料, 设有 g 层, 每层有 $n_k (k = 1, 2, \dots, g)$ 例观察对象, 其中有 d_k 个病例, 则似然函数为:

$$L = \prod_{k=1}^g P_k^{d_k} (1 - P_k)^{n_k - d_k} \quad (10-8)$$

式中, P_k 表示第 k 层阳性结果发生的概率。

2) 优势比估计

某一个因素两个不同水平 (c_1, c_0) 优势比的估计值为

$$\hat{OR}_i = \exp[b_i(c_1 - c_0)] \tag{10-9}$$

OR_i 的置信区间可以利用 b_i 的抽样分布来估计, 大样本情况下, 服从近似的正态分布。特殊情况下, 若自变量 X_i 只有暴露和非暴露两个水平, 则优势比 OR_i 的 $(1 - \alpha)$ 置信区间为:

$$\exp(b_i \pm u_{\alpha/2} S_{b_i}) \tag{10-10}$$

2. 假设检验

如同多重线性回归的假设检验, 与 logistic 回归方程有关的假设检验也包括两个内容: 一是整个模型的检验, 也就是检验因变量与自变量的关系是否能用所建立的回归方程表示; 二是单个回归系数的检验, 也就是检验单个自变量对于因变量的影响是否存在。

当对模型中的回归系数是否均为零做检验, 检验假设为:

$$H_0: \beta_0 = \beta_1 = \cdots = \beta_p = 0 \quad H_1: \beta_0, \beta_1, \cdots, \beta_p \text{ 不全为零}$$

对某一个回归系数进行检验, 检验假设为:

$$H_0: \beta_i = 0 \quad H_1: \beta_i \neq 0$$

常用的检验方法有似然比检验、Wald χ^2 检验和计分检验。

1) 似然比检验(likelihood ratio test)

似然比检验是比较两种不同条件下的对数似然函数值, 检验统计量是对数似然函数中两个包含不同参数个数的模型的对数似然函数之差。既适合单个自变量的假设检验又适合多个自变量的同时检验。具体做法是, 通过拟合一个不包含准备检验的变量的 logistic 回归模型, 它的对数似然函数值为 $\ln L_0$, 把要检验的变量纳入, 得到 $\ln L_1$ 。似然比统计量 G 的计算公式为:

$$G = 2(\ln L_1 - \ln L_0) \tag{10-11}$$

当样本量较大时, 零假设下得到的 G 统计量近似服从自由度为 $d(d = p - 1)$ 的 χ^2 分布, 若只对一个回归系数检验, 则 $d = 1$ 。

2) Wald χ^2 检验

Wald χ^2 检验比较适合单个变量的检验, 将 β_i 的估计值 b_i 与 0 比较, 并用标准误 S_{b_i} 作为参照。检验假设

$$\begin{aligned} H_0: \beta_i &= 0 \\ H_1: \beta_i &\neq 0 \end{aligned}$$

统计量

$$\chi^2 = \left(\frac{b_i}{S_{b_i}} \right)^2 \tag{10-12}$$

χ^2 服从自由度为 1 的 χ^2 分布。

对于大样本, 统计量

$$u = \frac{b_i}{S_{b_i}} \tag{10-13}$$

近似服从正态分布。

3) 计分检验(score test)

计分检验的检验假设

$$H_0: \beta_i = 0; H_1: \beta_i \neq 0$$

统计量

$$\text{score} = S'(\text{COV})S \tag{10-14}$$

其中, S 是对数似然函数的一阶偏导数值, S' 是列向量 S 的转置, COV 是模型的方差协方差矩阵。大样本下, 计分统计量服从自由度为模型中被估计参数个数的 χ^2 分布。

上述三种参数检验的方法中, 似然比检验是将含有不同自变量的模型之间进行比较, 既可以用于单个自变量的假设检验又适合多个自变量同时检验, 结果相对比较可靠。Wald 检验比较适合单个自变量的检验, 但是结果偏于保守。计分检验与传统 Mantel-Haenszel 分层检验方法非常接近, 并且在小样本情况下, 计分检验统计量的分布可能更接近 χ^2 分布, 所以导致 I 类错误的可能性要小些。大样本情况下, 三种检验方法所得的结果是一致的。

3. 样本量的估计

经验估计样本量为协变量个数的 10 ~ 15 倍。1998 年, F. Y. Hsieh 等提出了以下样本量估计公式。

1) 一个自变量的情况

当该自变量为正态分布的连续变量时, 样本量为

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2}{P_1(1 - P_1)\ln(\text{OR})^2} \quad (10-15)$$

式中, $z_{\alpha/2}$ 为标准正态分布双侧 α 临界值, z_{β} 为标准正态分布单侧 β 临界值, P_1 为自变量取均值条件下事件发生的概率, $\ln(\text{OR})$ 为能够检测到的效应大小。

当该自变量为二值变量时, 样本量为

$$n = \frac{\{z_{\alpha/2} [P/(1 - P)/B]^{1/2} + z_{\beta} [P_1/(1 - P_1) + P_2/(1 - P_2)(1 - B)/B]^{1/2}\}^2}{(P_1 - P_2)^2(1 - B)} \quad (10-16)$$

其中, B 为 $X=1$ 所对应的个体数占总个体数的比例; P_1 为 $X=0$ 所对应样本的事件发生频率; P_2 为 $X=1$ 所对应样本的事件发生概率。 P 为总的事件发生率, $P = (1 - B)P_1 + BP_2$ 。

2) 多个自变量的情况

首先从所有的自变量中确定一个主要感兴趣的自变量(假定为 X_1), 根据这个自变量是正态还是二值变量, 选择式(10-15)和式(10-16)估算 n , 则总的样本含量 N 为:

$$N = \frac{n}{1 - \rho_{112, 3, \dots, p}^2} = \text{VIP}_p \quad (10-17)$$

式中, $\rho_{112, 3, \dots, p}$ 为以 X_1 为因变量关于自变量 X_2, \dots, X_p 做多重回归所得的决定系数。 $(1 - \rho_{112, 3, \dots, p}^2)^{-1}$ 称为方差膨胀系数。

4. 回归模型中自变量的筛选

多重 logistic 回归的变量筛选与多重线性回归一样, 为了使得建立的 logistic 比较稳定和便于解释, 会尽可能筛选出使回归效果显著的自变量进入模型。和多重线性回归极为相似, logistic 自变量筛选有前进法(FORWARD)、后退法(BACKWARD)和逐步法(STEPWISE), 另外, 还有最优子集法(SCORE)。逐步 logistic 回归和逐步线性回归分析过程极为相似, 但是检验统计量不再是 F 检验统计量, 而是用似然比统计量、Wald 统计量和计分统计量之一来决定变量是引入还是从模型中剔除。

最优子集法常与 $\text{BEST} = n$ 连在一起, BEST 用于显示具有最优子集的自变量的个数。在统计分析的基础上, 结合专业知识, 从可解释性、简约性、变量的易得性等方面综合考虑, 选择“最佳模型”, 通常需要不断调整并最后确定。

5. logistic 回归模型的拟合优度检验

logistic 回归模型的拟合优度是通过比较模型预测结果与实际观测时间发生与不发生的频数有

无差别来进行检验的。预测结果与实际观测越吻合,说明模型的拟合效果越好。假设检验中, H_0 为模型的拟合效果好, α 可以取 0.1 或 0.2。常用的方法有以下几种。

1) 偏差 (deviance) 检验

统计量计算公式为

$$\chi^2_D = 2 \sum_{i=1}^M \sum_{j=1}^K O_{ij} \ln \left(\frac{O_{ij}}{n_i \hat{p}_j} \right) \tag{10-18}$$

M 是自变量不同取值的组合数, K 是反应变量的分类数, O_{ij}, \hat{P}_{ij} 是第 i 中组合与第 j 个分类下实际频数和拟合概率, $n_i = \sum_{j=1}^k O_{ij}$ 。自由度为 (MK-估计的参数个数)。

2) pearson χ^2 检验

统计量计算公式为

$$\chi^2_P = 2 \sum_{i=1}^M \sum_{j=1}^K \frac{(O_{ij} - n_i \hat{p}_j)^2}{n_i \hat{p}_j} \tag{10-19}$$

自由度也为 (MK-估计的参数个数)。

多数情况下, χ^2_D 和 χ^2_P 数值接近,能得到相同的结论。然而,样本量较小时,两种统计量的估计偏差较大,一般认为偏差统计量比 Pearson χ^2 统计量更优。

上面两种检验方法对样本量的要求:

- (1) 各自变量取值在不同组合情况下至少有 10 例观测;
- (2) 80% 自变量取值的不同组合要有 5 个以上预测事件数;
- (3) 自变量组合下预测时间数应该大于 2。

3) Homser-Lemeshow 统计量

当自变量尤其是连续型自变量纳入模型的数量增多时,变量之间不同取值的组合数量会很大,每组只有很少观测,偏差检验和 Pearson χ^2 检验的自由度增大,结果变得不可靠,这时可以采用 Homser-Lemeshow 统计量来评价模型的拟合优度。Homser-Lemeshow 统计量是一种类似于 Pearson χ^2 的统计量。统计量计算公式为

$$\chi^2_{HL} = \sum_{g=1}^G \frac{(O_g - n_g \hat{p}_g)^2}{n_g \hat{p}_g (1 - \hat{p}_g)} \tag{10-20}$$

其中, G 代表分组数,且 $G \leq 10$; O_g 为第 g 组事件的实际数; n_g 为第 g 组中样本例数; \hat{p}_g 为第 g 组的预测事件概率; $n_g \hat{p}_g$ 为事件的预测数。自由度为 $G - 2$ 。

10.1.3 logistic 回归模型的流行病学意义

1. 参数 β_0 的意义

β_0 表示当各相关因素暴露剂量为 0 时个体发病与不发病概率之比的自然对数。

2. 参数 β_i 的意义

β_i 表示自变量 X_i 改变一个单位时 $\text{logit}(P)$ 的改变量。对比某一因素 X_i 的两种不同的暴露水平 $X_i = e_1, X_i = e_0$ 的发病情况,假定其他变量取值不变,则

$$\ln \text{OR}_i = \ln \left[\frac{P_1 / (1 - P_1)}{P_0 / (1 - P_0)} \right] = \text{logit}(P_1) - \text{logit}(P_0) = \beta_i (e_1 - e_0)$$

即

$$OR_i = \exp[\beta_i(e_1 - e_0)] \quad (10-21)$$

式中, P_1 和 P_0 分别表示 X_i 取值为 e_1 和 e_0 时的发病概率, OR_i 称为多变量调整后的优势比 (adjusted odds ratio), 表示扣除了其他自变量影响后所关注的该危险因素的作用。

特殊地, 如果 X_i 仅取

$$X_i = \begin{cases} 1 & \text{暴露} \\ 0 & \text{非暴露} \end{cases}$$

则暴露组与非暴露组出现阳性结果的优势比为

$$OR_i = \exp \beta_i \quad (10-22)$$

当 $\beta_i = 0$ 时, $OR_i = 1$, 说明自变量 X_i 对是否出现阳性结果不存在影响; 当 $\beta_i \neq 0$ 时, $OR_i \neq 1$, 说明自变量 X_i 可能是危险因素或是保护因子。如果 $OR_i > 1$, 说明自变量 X_i 是危险因素, $OR_i < 1$, 则说明自变量 X_i 是保护因素。

由于 logistic 回归模型中参数与优势比 OR 的对应关系, 使得 logistic 回归分析常被用于疾病相关因素的分析。由于 OR_i 值与模型中的常数项无关, β_0 在危险因素分析中被视为无效参数。在研究像恶性肿瘤这种发病率较低 (例如 $\leq 10\%$) 的慢性疾病时, 因为 P 很小, 优势比可以作为相对危险度 (relation risk, RR) 的近似估计, 即

$$OR = \frac{P_1/(1 - P_1)}{P_0/(1 - P_0)} \approx \frac{P_1}{P_0} = RR \quad (10-23)$$

也就是说得到某一因素的回归系数估计值以后, 可以得到危险因素不同水平下相对危险度的估计值。

logistic 回归分析是假定 logistic 回归模型是自变量线性组合的函数, 若这种组合非线性或者存在交互, 将会影响模型拟合。

10.2 条件 logistic 回归

10.2.1 基本原理

在流行病学的病例 - 对照研究中, 为了防止某些重要非试验因素成为混杂因素 (即非试验因素在试验因素不同水平组中的影响不均衡), 常把病例与对照按照年龄、性别等条件进行配对, 形成多个匹配组 (每一个匹配组可视为一个层)。匹配组中病例与对照人数比例是任意的, 最常用的是 $1:M$ 配对研究 (一般 $M \geq 3$)。

配对资料能控制影响因素的干扰, 用 P_i 表示第 i 层在一组危险因素作用下发病的概率, logistic 模型可以表示为:

$$P_i = \frac{1}{1 + \exp[-(\beta_{0i} + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m)]} \quad i = 1, 2, \cdots, n \quad (10-24)$$

β_{0i} 表示各层的效应, $\beta_1, \beta_2, \cdots, \beta_m$ 为待估计的参数。

与非条件 logistic 回归分析不同, 对于不同的匹配组 β_{0i} 各不相同, 但假定每个危险因素的致病能力在不同匹配组中相同。参数有 $\beta_{01}, \beta_{02}, \cdots, \beta_{0n}, \beta_1, \beta_2, \cdots, \beta_m$, 而实际上研究者并不关心配对因素, 可以不用估计 $\beta_{01}, \beta_{02}, \cdots, \beta_{0n}$, 而同一协变量在不同配对组中的作用是相同的, 条件 logistic 回归采用条件似然函数替代了一般 logistic 回归分析中的似然函数。研究表明, 样本量 > 200 时, 可以不考虑参数估计的偏差。使用条件 logistic 回归模型时, 配对组数不应 < 50 。若协变量较多, 组数还应增加, 或增大对照的数量。

10.2.2 参数估计与假设检验

条件 logistic 回归采用条件似然函数进行参数估计，方法与非条件 logistic 回归相同。用 $X_{it} = (X_{it1}, X_{it2}, \dots, X_{itm})$ 表示第 i 层内第 t 个观察对象危险因素的观察值，考虑第 i 个匹配组中的 $M + 1$ 个观察对象有 1 名病例的条件下，恰好第一个观察对象属于病例组的条件概率为：

$$\begin{aligned} L_i &= \frac{P(X_{i0}Y = 1) \prod_{t=1}^M P(X_{it}Y = 0)}{\sum_{t=0}^M [P(X_{it}Y = 1) \prod_{t=0, t \neq t}^M P(X_{it}Y = 0)]} \\ &= \frac{1}{1 + \sum_{t=1}^M \exp[\sum_{j=1}^m \beta_j (X_{ij} - X_{ioj})]} \end{aligned}$$

综合 n 个匹配组的条件似然函数为：

$$L = \prod_{i=1}^n \frac{1}{1 + \sum_{t=1}^M \exp[\sum_{j=1}^m \beta_j (X_{ij} - X_{ioj})]} \tag{10-25}$$

可以看到，匹配组效应的常数项被自动消去。采取与非条件 logistic 回归分析相同的方法，即用 Newton-Raphson 迭代方法求参数的估计值 $b_j (j = 1, 2, \dots, m)$ 及其标准误 S_{b_j} 。
假设检验的方法与非条件 logistic 回归分析相同，可参照前面内容进行。

10.3 logistic 回归分析的应用及其注意事项

10.3.1 logistic 回归分析的应用

1. 控制和校正重要非试验因素

生物医学中，观察对象的某一结局常常受到诸多因素的影响，这里既有试验因素也有重要非试验因素。控制重要非试验因素主要从两个方面进行：一是设计阶段通过分层、匹配设计等手段使非试验因素在各试验组之间均衡；二是统计分析时，若非试验因素较少，可以使用 Mantel-Haenszel 进行分层分析。但如果非试验因素较多，则分层较多，常会出现部分格子中频数为零的情况，使得上述解决方法得到的结论不稳定。logistic 回归技术，将所有的试验因素、非试验因素以及交互作用均体现在模型中，既充分利用了资料信息，又达到了在有效控制了非试验因素影响的条件下，研究解决因变量与试验因素间的依存关系的目的。

2. 筛选危险因素，做病因学分析

由 logistic 回归系数与 OR 的关系可见，logistic 回归模型可以用来对疾病的影响因素进行多因素分析。logistic 回归分析可用于病因探索与分析的病例 - 对照研究、队列研究以及横断面研究。在设计阶段，结合理论基础和专业对有可能影响结局的变量进行选择。由于可供考虑的变量太多，需要根据事先设定的检验水准对没有统计学意义的因素进行剔除，保证模型的相对较优。

3. 预测与判断

logistic 回归模型是一个概率型模型，因此可以用来预测某事件发生的概率 P 。临床上，可以根据患者的一些检查指标，判断事件（如发病）发生的概率到底有多大。

4. 半数数量分析

在一些药物或毒物的剂量-反应研究中,由于同种类动物个体间的差异,每一只动物的药物耐受量会有所不同。恰有一半动物起“阳性反应”的剂量称为半数数量。若用 P 表示剂量为 X 时的阳性率,则有以下式:

$$P = \frac{1}{1 + \exp[-(\beta_0 + \beta \ln X)]} \quad (10-26)$$

若 β_0 和 β 的样本估计为 b_0 和 b , 对数半数数量为:

$$\ln X_{50} = \frac{b_0}{b} \quad (10-27)$$

方差为:

$$D(\ln X_{50}) = \left(-\frac{b_0}{b}\right)^2 \left\{ \frac{D(b_0)}{b_0^2} - \frac{2D(b_0, b)}{b_0 b} + \frac{D(b)}{b^2} \right\} \quad (10-28)$$

对数半数数量的 95% 的置信区间为:

$$-\frac{b_0}{b} \pm u_{\alpha/2} \sqrt{D(\ln X_{50})} \quad (10-29)$$

10.3.2 logistic 回归分析的注意事项

1. 样本量

logistic 回归分析的统计推断需要建立在大样本基础上, 所以为了得到可靠的参数估计, 需要足够的样本量。经验上, 病例和对照人数应至少各有 30 ~ 50 例, 随着自变量个数的增加, 自变量各水平的交叉分类数迅速增加, 各组的例数也应该相应增大。对于配对资料, 样本的匹配组数应为纳入方程中自变量个数 p 的 20 倍以上, 也就是 $n \geq 20p$ 。

2. 变量的赋值

logistic 回归分析要求研究对象之间彼此独立, 所以不适合有聚集性特征的资料。建立 logistic 回归模型的自变量可以是无序分类变量、有序分类变量和数值变量。二分类变量一般为 0 和 1; 有序多分类变量按等级秩次或哑变量赋值; 无序多分类变量应该转化成哑变量形式。对于数值型连续变量, 是否向分类变量转化以及何时转化需要结合专业解释进行综合考虑。

即便对于统一资料分析, 变量采用了不同的取值形式, 参数的含义、量值以及符号都可能发生变化。例如, 在做危险因素分析时, 若自变量是一个定量指标, 最好能够按取值大小分成几组, 变成等级资料或使用哑变量来进行分析, 以利于结果的合并和解释。并且, 哑变量的赋值方法不同, 结果的解释也不相同, 有的赋值编码更利于参数解释, 有的赋值编码更利于假设检验。

3. 模型拟合优度检验

模型的假设检验只能说明建立的模型是否具有统计学意义, 不能描述模型的拟合效果。拟合优度检验能够评价模型预测值和实际观测值的一致性。拟合优度高说明自变量的预测能力高。常用的模型拟合优度检验方法有三种。

1) 偏差 (deviance)

$$D = 2(\ln L^* - \ln L) = 2 \sum_{k=1}^g \left[r_k \ln \left(\frac{r_k}{n_k \hat{P}_k} \right) + (n_k - r_k) \ln \left(\frac{n_k - r_k}{n_k - n_k \hat{P}_k} \right) \right] \quad (10-30)$$

其中, $n_k, r_k, n_k - r_k$ 分别是各层的观察例数、阳性数和阴性数; \hat{P}_k 表示各层概率的估计值; g 为层数。

L^* 表示模型与实际资料拟合情况下对数似然值, L 为待检验模型的最大似然值。 D 服从自由度为 $\nu = g - p - 1$ 的 χ^2 分布, 表示样本数据与 logistic 模型的拟合程度, 当 $D \approx \chi^2 > \chi^2_{\alpha, \nu}$ 时, 说明数据和模型拟合不好。

2) Pearson χ^2

$$\chi^2 = \sum_{k=1}^g \frac{(r_k - n_k \hat{P}_k)^2}{n_k \hat{P}_k (1 - \hat{P}_k)} \quad (10-31)$$

其中, $n_k, r_k, n_k - r_k$ 分别是各层的观察例数、阳性数和阴性数; \hat{P}_k 表示各层概率的估计值; g 为层数。

大样本下, 这两种拟合检验方法结果接近, 均服从自由度为 $\nu = g - p - 1$ 的 χ^2 分布 (g 为层数, p 为模型中自变量的个数)。如果分层过多, 一些层内的观测可能过少, 计算得到的统计量可能偏离 χ^2 分布, 这时, 用偏差与自由度进行比较, 若 $D > \nu$, 则表示拟合不够理想。

3) 广义决定系数 (generalized coefficient of determination)

广义决定系数是一种类似于多重线性回归中的决定系数的综合性指标。

$$R^2 = 1 - \left\{ \frac{L_0}{L_1} \right\}^{\frac{2}{n}} \quad 0 \leq R^2 < 1 \quad (10-32)$$

其中, L_0 为模型中不包含任何自变量的似然函数; L_1 为模型中包含所有自变量的似然函数值。 R^2 值越接近 1, 说明模型的拟合效果越好, 最大值为 $R^2_{\max} = 1 - \left\{ L_0 \right\}^{\frac{2}{n}}$ 。为了使得决定系数的理论上能够等于 1, 常使用最大调整系数

$$R^2_{\text{res}} = \frac{R^2}{R^2_{\max}} \quad 0 \leq R^2_{\text{res}} \leq 1 \quad (10-33)$$

4. 有序 logistic 回归和多分类 logistic 回归

当因变量呈等级或程度差别的资料时, 应用有序 logistic 回归 (ordinal logistic regression), 基于累计概率构建回归模型。若因变量 Y 包括 g 个等级 (Y 取值为 $1, 2, \dots, g$), 影响 Y 取值的 m 个自变量 X_1, X_2, \dots, X_m , 则有序 logistic 回归模型为

$$\ln \left(\frac{P(Y \leq j)}{1 - P(Y \leq j)} \right) = \beta_{0j} + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \quad (10-34)$$

其中, $j = 1, 2, \dots, g - 1$ 。因变量 Y 包括 g 个类别, 有序 logistic 回归包括 $(g - 1)$ 个方程, β_{0j} 为第 j 个回归方程的常数项, $\beta_1, \beta_2, \dots, \beta_m$ 为第 j 个回归方程自变量 X_1, X_2, \dots, X_m 的回归系数。有序 logistic 回归中 $(g - 1)$ 个方程中自变量回归系数相同, 不同类别累计概率的差别体现在常数项之上。自变量 X 改变一个单位的优势比 OR 可以使用回归系数获得。流行病学上有序 logistic 回归因变量 Y 赋值时, 专业上最不利的等级赋予最小值, 最有利的等级赋予最大值。

当因变量是无序多分类变量 (multinomial logistic regression) 模型进行分析时, 选择的是因变量 Y 中众多类别之一为参照, 拟合的是其余类别相对于参照类别的 logistic 回归模型。若因变量 Y 包括 g 个分类 (Y 取值为 $1, 2, \dots, g$), 另有影响 Y 取值的 m 个自变量 X_1, X_2, \dots, X_m 则多分类 logistic 回归模型可表示为

$$\ln \left(\frac{P(Y = j)}{P(Y = g)} \right) = \beta_{0j} + \beta_{1j} X_1 + \beta_{2j} X_2 + \dots + \beta_{mj} X_m \quad (10-35)$$

其中, $j = 1, 2, \dots, g - 1$ 。应变量 Y 包括 g 个类别, 多分类 logistic 回归包括 $(g - 1)$ 个方程, β_{0j} 为第 j 个回归方程的常数项, $\beta_{1j}, \beta_{2j}, \dots, \beta_{mj}$ 为第 j 个回归方程自变量 X_1, X_2, \dots, X_m 的回归系数。多分

类 logistic 回归拟合的剩余各类别相对于参照类别的回归方程中的自变量回归系数可以不同。自变量 X 改变一个单位的优势比 OR 可以使用回归系数获得。

有序 logistic 回归模型和多分类 logistic 回归模型对参数估计和假设检验的方法与二分类 logistic 回归模型类似。

10.4 案例分析

【例 10-1】 在研讨肾细胞癌转移的有关临床病理因素研究中，收集了一批进行根治性肾切除术患者的肾癌标本资料，现从中抽取 26 例资料进行 logistic 回归分析实例。因素变量表示及赋值情况如表 10-1 所示，具体数据如表 10-2 所示。

表 10-1 肾细胞癌 5 个相关因素及赋值

因 素	变量名	赋值说明
年龄	X_1	连续变量
肾细胞癌血管内皮生长因子(VEGF)	X_2	阳性表述从低到高分 3 级
肾细胞癌组织内微血管数(MVC)	X_3	连续变量
肾癌细胞核组织学分级	X_4	由低到高分 4 级
肾细胞癌分期	X_5	由低到高分 4 级
肾细胞癌转移情况	Y	无转移 =0, 有转移 =1

表 10-2 肾细胞癌相关因素的病例对照研究

序号	X_1	X_2	X_3	X_4	X_5	Y
1	59	2	43.4	2	1	0
2	36	1	57.2	1	1	0
.....						
25	58	1	128.0	4	3	0
26	60	3	19.8	4	3	1

SAS 程序如下：

```
data nkkysj10_1;      * 建立数据集;
input x1 -x5 y@@ ;
output;
cards;
59 2 43.4 2 1 0
36 1 57.2 1 1 0
61 2 190.0 2 1 0
58 3 128.0 4 3 1
55 3 80.0 3 4 1
61 1 94.4 2 1 0
38 1 76 1 1 0
42 1 240.0 3 2 0
50 1 74.0 1 1 0
58 3 68.6 2 2 0
68 3 132.8 4 2 0
25 2 94.6 4 3 1
52 1 56.0 1 1 0
31 1 47.8 2 1 0
36 3 31.6 3 1 1
```



```
42 1 66.2 2 1 0
14 3 138.6 3 3 1
32 1 114.0 2 3 0
35 1 40.2 2 1 0
70 3 177.2 4 3 1
65 2 51.6 4 4 1
45 5 124.0 2 4 0
68 3 127.2 3 3 1
31 2 124.8 2 3 0
58 1 128.0 4 3 0
60 3 19.8 4 3 1
;
run;
ods html;
proc logistic descending;          * 进行 logistic 回归分析——前进法;
model y=x1 -x5/selection=forward;
run;
proc logistic descending;          * 进行 logistic 回归分析——后退法;
model y=x1 -x5/selection=backward;
run;
proc logistic descending;          * 进行 logistic 回归分析——逐步回归法;
model y=x1 -x5/selection=stepwise  sle=0.10 sls=0.05;
run;
ods html close;
quit;
```

程序说明：proc logistic 语句中加上选项 descending，是为了按照结果变量取值降序{y = 1(有转移)}的概率拟合模型。如果不加这个选项，则会默认以 y = 0(无转移)的概率进行模型拟合，这时，因变量的排序水平发生颠倒，且所有参数估计的符号相反，OR 为原来的倒数。模型定义 y 为因变量，x1 - x5 为自变量，selection = stepwise 为逐步回归法筛选自变量，sle = 0.10, sls = 0.05 表示制定变量入选和剔除模型的显著性水平分别为 0.10 和 0.05。

SAS 运行结果：

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	35.542	21.351
SC	36.800	25.125
-2 Log L	33.542	15.351

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	18.1905	2	0.0001
Score	13.5580	2	0.0011
Wald	5.6595	2	0.0590

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
3.5760	3	0.3110

以上是前进法筛选自变量的步骤，自变量 X_2 和 X_4 被选入模型，模型的收敛结果满意，显示了模型拟合统计量 Model Fit Statistics，给出了不同拟合情况下模型的似然函数的三种判断标准 AIC、SC 和 -2LOG ，还显示了无效假设为 $\beta = 0$ 的假设检验结果，分别采用了 Likelihood Ratio、Score 和 Wald 三种方法，结果显示前两种检验有统计学意义，Wald 检验无统计学意义。最后给出了残差 χ^2 的检验结果，没有统计学意义。

前进法中，最终进入回归方程的变量有 X_2 和 X_4 。

Summary of Forward Selection					
Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq
1	x4	1	1	12.0916	0.0005
2	x2	1	2	4.2134	0.0401

最大似然估计计算进入回归方程的 2 个变量的参数估计。2 个回归系数均为正数，且检验结果 $P < 0.05$ ，可以认为 2 个因素均为肾细胞癌转移的危险因素。 X_2 对应的 OR 为 3.248，表示肾细胞癌血管内皮生长因子(VEGF)的阳性表述每增加一个级别，肾癌细胞转移的危险度增加 3.248 倍。肾癌细胞核组织学分级等级增加一级，肾癌细胞转移的危险度增加 10.210 倍。由此可以得到 logistic 回归方程为

$$\ln\left(\frac{P}{1-P}\right) = -10.2901 + 1.1779X_2 + 2.3234X_4$$

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-10.2901	4.4718	5.2952	0.0214
x2	1	1.1779	0.6942	2.8789	0.0897
x4	1	2.3234	1.0107	5.2839	0.0215

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
x2	3.248	0.833	12.661
x4	10.210	1.408	74.023

预测概率与观察到结果的关联性，它包括对不同结果的个数对 4 种秩相关指数的分析。

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	91.5	Somers' D	0.850
Percent Discordant	6.5	Gamma	0.867
Percent Tied	2.0	Tau-a	0.400
Pairs	153	c	0.925

后退法只进入了 X_4 。

Summary of Backward Elimination					
Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	x5	1	4	0.8230	0.3643
2	x1	1	3	1.0633	0.3025
3	x3	1	2	1.2304	0.2673
4	x2	1	1	2.8789	0.0897

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.4600	2.2721	8.0838	0.0045
x4	1	2.0521	0.7322	7.8546	0.0051

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
x4	7.785	1.853	32.697

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	85.6	Somers' D	0.804
Percent Discordant	5.2	Gamma	0.885
Percent Tied	9.2	Tau-a	0.378
Pairs	153	c	0.902

逐步回归只有 X_4 进入回归方程。

Summary of Stepwise Selection							
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
	Entered	Removed					
1	x4		1	1	12.0916		0.0005
2	x2		1	2	4.2134		0.0401
3		x2	1	1		2.8789	0.0897

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.4600	2.2721	8.0838	0.0045
x4	1	2.0521	0.7322	7.8546	0.0051

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
x4	7.785	1.853	32.697

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	85.6	Somers' D	0.804
Percent Discordant	5.2	Gamma	0.885
Percent Tied	9.2	Tau-a	0.378
Pairs	153	c	0.902

结合 3 种自变量筛选的方法，回归方程中只进入了 X_4 ， $\ln\left(\frac{P}{1-P}\right) = -6.4600 + 2.0521X_4$

【例 10-2】 研究肥胖(X_1 , 肥胖为 1, 不肥胖为 0)、口服避孕药雌激素(X_2 , 用药为 1, 不用药为 0)与子宫内膜癌(Y , 病例为 0, 对照为 1)的关系, 采用 1:2 配对做病例 – 对照研究, 共调查 20 个配比组(见表 10-3)。试分析肥胖、口服避孕药雌激素与子宫内膜癌的关系。

表 10-3 肥胖、口服避孕药雌激素与子宫内膜癌的关系

配比组	肥胖	口服避孕药雌激素	子宫内膜癌	配比组	肥胖	口服避孕药雌激素	子宫内膜癌	配比组	肥胖	口服避孕药雌激素	子宫内膜癌
1	1	1	0	1	0	0	1	1	0	0	1
2	1	1	0	2	1	1	1	2	0	1	1
.....											
19	1	0	0	19	0	1	1	19	0	1	1
20	1	1	0	20	0	1	1	20	0	0	1

SAS 程序如下：

```
Data nkkysj10_2;
input  idx1 x2 y@@ ;
cards;
1  1  1  0  1  0  0  1  1  0  0  1
2  1  1  0  2  1  1  1  2  0  1  1
3  1  1  0  3  0  1  1  3  1  1  1
4  0  1  0  4  0  0  1  4  0  1  1
5  0  0  0  5  1  0  1  5  0  1  1
6  1  1  0  6  0  0  1  6  1  0  1
7  1  1  0  7  0  1  1  7  1  1  1
8  1  1  0  8  1  0  1  8  1  1  1
9  1  0  0  9  1  1  1  9  1  1  1
10 0  1  0 10 0  1  1 10 0  0  1
11 0  1  0 11 0  1  1 11 1  0  1
12 0  1  0 12 0  1  1 12 0  1  1
13 1  1  0 13 0  0  1 13 1  1  1
14 1  1  0 14 0  0  1 14 1  0  1
15 1  1  0 15 0  0  1 15 0  1  1
16 0  1  0 16 0  1  1 16 0  1  1
17 0  1  0 17 0  0  1 17 1  0  1
18 1  1  0 18 1  0  1 18 0  1  1
19 1  0  0 19 0  1  1 19 0  1  1
20 1  1  0 20 0  1  1 20 0  0  1
;
run;
ods html;
proc phreg;
model y=x1 x2/
risklimits ties=discrete;
strata id;
run;
ods html close;
```

调用 phreg 过程步，指定配对 logistic 回归分析，分层变量为 id。

SAS 运行结果：

Analysis of Maximum Likelihood Estimates								
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
x1	1	1.82388	0.84948	4.6098	0.0318	6.196	1.172	32.748
x2	1	1.58959	0.80897	3.8611	0.0494	4.902	1.004	23.930

从结果可以看出，肥胖、口服避孕药雌激素与子宫内膜癌的关系有统计学意义，肥胖者患子宫内膜癌的危险性约为非肥胖者的 6 倍 ($OR\hat{=} 6.196$)；口服避孕药雌激素患子宫内膜癌的危险性约为不口服避孕药雌激素者的 5 倍 ($OR\hat{=} 4.902$)。

(孔雨佳 王素珍 石福艳)

第11章 生存分析

11.1 生存分析中的基本概念

临床试验研究中,经常会对各个观察对象进行随访观察,记录每个时点上事件的发生情况,用以评价疗效。在比较两种疗法的疗效时,不同的治疗药物产生疗效的时间不一样,不能单纯用见效来笼统描述和分析。生存分析(survival analysis)是将事件发生的结果和随访时间两者结合在一起的统计分析方法,可以用来描述和比较生存过程,以及分析影响生存时间的因素。生存分析又称事件时间分析、事件历史分析或失效时间分析,不仅在医学上,而且在经济学、社会学以及工程学上都有广泛的应用。

大多数医学研究中,虽然研究期限是固定的,但是观察对象在此期间先后进入研究,随访直至终点事件发生或研究结束,我们用生存时间来描述从观察开始到特定终点事件出现经历的时间长度。临床试验及其随访资料的特点是:一部分研究对象可观察到结局(如死亡),从而得到准确的生存时间,这种资料提供的信息是完全的,称为完全数据(complete data);另一部分数据,由于中途失访或到观察结束时仍存活,导致无法获取准确的生存时间,无法提供完整信息,称为不完全数据,或者截尾数据(censored data)。生存分析方法不仅可以分析完全数据的资料,也可以分析不完全数据的资料。

11.1.1 基本概念

1. 生存时间

起始事件(initial event)是反映研究对象生存过程的起始特征的事件。终点事件(failure event)又称死亡事件、失效事件(failure event),是指研究者所关心的研究对象的特定结局,反映治疗效果特征的事件。起始事件、终点事件、时间单位等应该在研究设计阶段确定。

生存时间(survival time)又称失效时间(failure time)是任何两个有联系的起始事件和终点事件间隔,记为随机变量 T , $T \geq 0$, 常用符号 t 表示。狭义的生存时间定义为患某种疾病的病人从发病(或治疗结束)到死亡所经历的时间跨度,广义的生存时间定义为从某种事件起始到终止所经历的时间跨度,甚至可以为医疗费用或医疗设备使用次数等。生存时间通常不服从正态分布,而是呈偏态分布,如指数分布、Weibull 分布、Gompertz 分布、对数分布等。

2. 截尾值

个体的生存时间可以确切获得为完整数据(complete data),若得不到确切的生存时间,则称为不完全数据(incomplete data)。生存结局(status)分为“死亡”和“删失”两种。截尾值,又称删失值(censored value),指在随访过程中,由于某种原因未能观察到病人的明确结局(即终止事件),所提供的生存时间信息是不完整的。临床实践中很难获得全部研究对象生存时间的全部信息,产生截尾值的主要原因:病人失访;病人的生存期超过了研究的终止期;动物实验规定的期限内未出现需要的事件或动物丢失了等。删失可细分为左删失(left censored)、区间删失(interval censored)和右删失(right censored)三类。如果感兴趣的终点事件在知晓时间之前发生,则称为左删失;如果感兴

趣的终点事件在知晓事件之后发生，则称为右删失。删失值所占的比例不宜太大，并且删失的模式应该是随机的。绘制的年历时间和生存时间的图中，“x”表示“死亡”，“o”表示“删失”。

3. 生存率与风险函数

生存函数(survival function) 又称累积生存率(cumulative probability of survival), 简称生存率(survival rate)或生存函数(survival function), 记为 $S(t)$, 表示具有协变量 X 的观察对象的生存事件 T 大于 t 的概率, 记为 $S(t)$ 。

如果资料不存在删失数据, 则

$$S(t_k) = P(T \geq t) = \frac{t \text{ 时刻仍存活的例数}}{\text{观察总例数}} \tag{11-1}$$

如果资料含有删失数据, 用概率乘法定理将各时段生存概率相乘

$$S(t_k) = p_1 \cdot p_2 \cdots p_k = S(t_{k-1}) \cdot p_k \tag{11-2}$$

生存率的标准误为

$$SE(S(t_k)) = S(t_k) \sqrt{\sum_{j=1}^k \frac{q_j}{p_j n_{0j}}} \tag{11-3}$$

观察起点即 $t=0$ 时的生存率为 1; 当观察期无穷大时, 生存率为 0。
年生存概率 p = 该年活满一年的人口数/年初人口数 = $(n_0 - d) / n_0$ 。
若年内有删失, 分母用校正人口数: 校正人口数 = 年初人口数 - 删失例数/2 = $n_0 - c/2$ 。
生存概率指单个时段的概率, 而生存率指从 0 ~ t 多个时段的累积生存概率。

4. 风险函数(hazard function)

风险率(hazard rate) 又称风险函数(hazard function), 表示已经活过时间 t 的观察对象在时刻 t 的瞬时死亡率, 记为:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P[(t \leq T < t + \Delta t) (T \geq t)]}{\Delta t} = \frac{f(t)}{S(t)} \tag{11-4}$$

风险率为非负数, 可以大于 1。

累积风险函数(cumulative hazard function) 记为 $H(t)$, 与生存函数的关系为

$$H(t) = -\ln S(t) \tag{11-5}$$

5. 死亡率与死亡密度函数

死亡概率密度函数(probability density function, pdf), 表示具有协变量 X 的所有观察对象在 t 时刻的瞬时死亡率(instantaneous rate):

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t)}{\Delta t} = F'(t) \tag{11-6}$$

死亡概率(mortality rate), 或称累积分布函数(cumulative distribution function, cdf) 或累积死亡概率, 表示一个观察对象从开始观察到时间 t 为止的死亡概率, 指死于某时段内的可能性大小:

$$F(t) = P(T \leq t) \tag{11-7}$$

年生存概率表示年初尚存人口存活满一年的可能性:

$$\text{年生存概率} = \frac{\text{活过该年人数}}{\text{某年年初人口数}} \tag{11-8}$$

年死亡概率表示年初尚存人口在今后 1 年内死亡的可能性:

$$\text{年死亡概率} = \frac{\text{该年内死亡人数}}{\text{某年年初人口数}} \tag{11-9}$$

$$\text{年死亡率} = \frac{\text{该年内死亡人数}}{\text{某年平均人口数}} \quad (11-10)$$

若年内有删失, 分母用校正人口数。

生存函数、风险函数、死亡函数具有以下关系:

$$h(t) = \frac{f(t)}{S(t)} \quad (11-11)$$

11.1.2 资料收集

1. 随访内容

研究内容包括确定开始随访的时间、随访的结局以及终止随访的时间, 并记录影响生存时间的有关因素以便分析对生存时间的影响。随访的结局可能有以下几种: (1)“死亡”, 即处理失效, 终止时间为“死亡”时间; (2)生存但中途失访; (3)死于其他与疾病无关的原因; (4)随访结束时, 观察对象仍存活。

分组变量(处理方法)和其他的协变量(性别、年龄、职业、文化程度等)也是需要收集的内容。

生存资料要求: 死亡例数不宜太少, 否则会出现偏性; 截尾原因无偏性; 生存时间尽可能记录准确。

2. 随访方式

若全部对象同时接受处理, 随访以最后一例出现结果或规定的随访结束时间。若随访对象在不同时间接受处理, 完成一定数量随访病例后决定随访截止时间或按事先规定的随访截止时间截止随访。

3. 生存研究的主要内容

(1) 描述生存过程。

研究生存时间的分布特点, 估计生存率及平均存活时间, 绘制生存曲线等。根据生存时间的长短, 可以估计出各时点的生存率, 并根据生存率来估计中位生存时间, 同时也可根据生存曲线分析其生存特点。常使用 Kaplan-Meier 法(乘积极限法)进行分析, 对于频数表资料采用寿命表法。生存率的分析属于非参数统计方法。

(2) 比较生存过程。

通过生存率及其标准误对各样本的生存率进行比较, 以探讨各总体的生存过程是否有差别。不同生存率之间的比较一般使用 log-rank 检验与 Breslow 检验。

(3) 影响生存时间的因素的分析。

通过对生存分析模型探讨影响生存时间的因素, 一般用生存时间和结局作为因变量, 而将影响它们的因素作为自变量, 筛选影响生存时间的保护因素和危险因素。

11.1.3 生存分析的基本方法

1. 非参数法

非参数法是不考虑资料的分布, 只根据提供的顺序统计量对生存率进行估计, 常用的方法有乘积极限法(product-limited method)和寿命表法(life table)。对于两个或者多个生存率的比较, 无效假设是假定两组或多组总体生存时间分布相同, 而不对具体分布形式及参数进行推断。

2. 参数法

参数法是当资料服从某种特定参数分布, 需要根据已知分布特点对影响生存时间进行分析。

参数法通过估计参数得到生存率的估计值，常用的方法有指数分布法、Weibull 分布法、对数正态回归分析法和对数 logistic 回归分析法等。

表 11-1 参数分布中相关参数的公式

分布	$f(t)$	$h(t)$	$H(t)$	$S(t)$
指数	$\lambda e^{-\lambda t}$	λ	λt	$e^{-\lambda t}$
Weibull	$\lambda p t^{p-1} e^{-\lambda t^p}$	$\lambda p t^{p-1}$	λt^p	$e^{-\lambda t^p}$
Gompertz	$a e^{bt} e^{[-a/b(e^{bt}-1)]}$	$a e^{bt}$	$a/b(e^{bt}-1)$	$e^{[-a/b(e^{bt}-1)]}$
对数 logistic	$ab t^{b-1} / (1 + at^b)^2$	$ab t^{-1} / (1 + at^b)$	$\ln(1 + at^b)$	$1 / (1 + at^b)$

3. 半参数法

半参数法兼有参数法和非参数法的特点，主要用于分析影响生存时间和生存率的因素，常用的方法为 Cox 模型分析法。

11.2 生存率的估计与生存曲线

11.2.1 小样本资料生存率及其标准误的计算

1. 乘积极限法 (product-limited method)

生存率的计算：当随访病例数较少时，使用乘积极限法 (product-limited method)，又称 Kaplan-Meier 法，既适合于大样本也适合于小样本。这种方法不需要对病例进行分组，用条件概率及概率乘法的原理直接计算生存率。

乘积极限法的具体步骤如下。

(1) 将生存时间排序：所有生存时间由小到大混合排列，生存时间相同的只列出一个。若同一数据有删失和未删失的，都列出且删失数据在后。

(2) 计算不同时刻的死亡数 d_i 、删失数 c_i 、期初例数 n_i 。

(3) 计算不同时刻死亡概率和生存概率： $q_i = d_i/n_i$, $p_i = 1 - q_i$ 。

(4) 计算生存率： i 时刻以及之前各个时点的生存概率连乘 $\hat{S}(t_i) = \prod_{j=1}^i p_j$ 。

2. 生存率的标准误计算

生存率的标准误计算有两种方法，公式为：

$$SE(\hat{S}(t_i)) = \hat{S}(t_i) \sqrt{\sum_{j=1}^i \frac{d_j}{n_j(n_j - d_j)}}$$

(11-12)

$$SE(\hat{S}(t_i)) = \hat{S}(t_i) \sqrt{\sum_{j=1}^i \frac{1 - \hat{S}(t_j)}{n_j - d_j}}$$

(11-13)

小样本情况下，后面的公式计算的标准误较大，所以更为保守。

假定生存率近似服从正态分布，总体生存率的 $(1 - \alpha)$ 可信区间为

$$\hat{S}(t_i) \pm u_{\alpha/2} \cdot SE[\hat{S}(t_i)]$$

(11-14)

为了保证生存区间在 0 和 1 之间，生存率的可信区间可以使用下面公式

$$\exp\{-\exp[\ln[-\ln\hat{S}(t_i)]] \pm u_{\alpha/2} \cdot \frac{SE[\hat{S}(t_i)]}{|\ln\hat{S}(t_i)| \cdot \hat{S}(t_i)}\}$$

(11-15)

3. 生存曲线

除了使用统计量, 图示法是更加直观的统计描述手段。用生存时间做横坐标, 生存率做纵坐标, 就可以得到一个右连续的阶梯型曲线, 这就是生存曲线, 也可以叫 Kaplan-Meier 曲线(K-M 曲线)。下降速度快在图形上表现为坡度大、曲线陡峭, 意味着较低的生存率或较短的生存期。

生存曲线(survival curve)是指将各个时点的生存率连接在一起的曲线图或表, 是阶梯曲线。以生存时间为横轴, 生存率为纵轴绘制的用以描述其生存过程的曲线。通过生存曲线的高低可以比较不同治疗方式之间的生存过程。生存曲线图上需要标出删失数据点。

4. 半数生存期

半数生存期(median survival time)又称中位生存时间或生存时间中位数, 是刚好有 50% 个体存活期大于该数据, 是用来对生存分析进行概括性描述最常用的统计量, 是反映集中趋势的指标。中位生存期越长, 说明疾病的预后越好。

计算方法主要有图解法和线性内插法。若删失个数太多, 超过一半, 则无法估计中位生存期。图解法利用生存曲线图从纵轴生存率 0.5 处画一条相对横轴的平行线与生存率曲线相交, 这个交点为中位生存时间。这种方法在例数较少时的误差较大。线性内插法是寻找两个生存率 $S(t_{i-1})$ 和 $S(t_i)$, 使得 $S(t_{i-1}) > 0.5$, $S(t_i) < 0.5$, 使用 $(S(t_{i-1}) - S(t_i)) / (S(t_{i-1}) - 0.5)$ 计算中位生存时间。

不同的百分位数生存时间可以使用相同的方法得到。其中, 生存期的四分位数间距 $= T_{25} - T_{75}$, 是反映离散程度的指标。

11.2.2 大样本资料的生存分析

1. 生存率的计算

大样本生存资料可以用寿命表法(life-table method)计算生存率, 它是 Kaplan-Meier 法的近似方法(频数表法)。寿命表法只估计时段右断点的生存率, 省略了时段内的生存率的估计。原理是求出患者治疗后各时期的生存概率, 然后根据概率的乘法原理, 将各时期生存概率相乘, 得到从开始观察到各个时点的生存期, 对生存率或生存分布之间的差别进行假设检验。具体步骤如下。

(1) 区间划分: 将全部生存时间进行区间划分 $[t_{i-1}, t_i)$, 对每一个病例进行分组, 在区间 $[t_{i-1}, t_i)$ 的病例可能出现三种情况, 即继续生存到区间终点 t_i 、在区间内死亡、在区间内截尾。

(2) 死亡人数 d_i , 截尾人数 c_i , 期初观察人数 n'_i 。

(3) 校正观察人数 $n_i = n'_i - c_i/2$ 。

(4) 死亡概率 $q_i = \frac{d_i}{n_i}$, 生存概率 $p_i = 1 - q_i$ 。

(5) 生存率:

$$\hat{S}(t_i) = \prod_{j \leq i} p_j \quad (11-16)$$

(6) 生存率标准误:

$$SE = \hat{S}(t_i) \sqrt{\sum_{j \leq i} \frac{q_j}{p_j N_j}} \quad (11-17)$$

2. 生存率曲线

寿命表的生存曲线估计的是每个时间区间右端点的生存率, 因每个时间区间内生存率的变化规律未知, 所以图形采用的是折线连接。和 K-M 法生存曲线一样, 寿命表的生存曲线同样可以采用图解法和线性内插法计算。

11.3 生存曲线的假设检验

11.3.1 log-rank 检验

生存曲线的 log-rank 检验也称为 Cox-Mantel 检验, 适合于两组或多组生存率之间的比较, 它的基本思想是实际死亡数和期望死亡数之间的比较。

1. 两组生存率比较的 log-rank 检验步骤

将两组资料混合后统一排序, 计算各组的期望死亡率, 求各组的期望死亡率之和, 计算 χ^2 值, 通过查 χ^2 界值表得到结论: 当 $P < 0.05$ 时, 认为在 $\alpha = 0.05$ 水平拒绝 H_0 , 可以认为两组生存率有差别。

如果对于大样本资料的生存率进行比较, 可以将其整理成频数表的形式。

2. log-rank 检验应用注意事项

(1) 这种方法既适用于小样本, 也适用于大样本以及多组比较。

(2) 实际死亡总数 A 与理论死亡总数 T 之比称为相对死亡比 (relative death ratio), $R = \frac{A}{T}$, 则相对危险度 (relative risk, RR) 估计值为两组相对死亡比之比。

(3) log-rank 检验用于整条生存曲线的比较。若比较两条生存曲线某个时间点处的生存率, 可以按照两个率比较的正态近似法, 公式为:

$$Z = \frac{\hat{S}_1(t) - \hat{S}_2(t)}{\sqrt{\text{SE}^2[\hat{S}_1(t)] + \text{SE}^2[\hat{S}_2(t)]}} \quad (11-18)$$

(4) log-rank 检验属于单因素分析方法, 应用时除了比较因素之外, 影响生存率的各协变量在两组间应该均衡可比, 表现在图上就是各组的生存曲线不能交叉, 可以使用分层或者多因素的方法来进行校正和均衡掉混杂因素。另外, 当假设检验的推断有差别的时候, 可以通过生存曲线、半数生存期以及相对危险度等指标来进行评价。

11.3.2 趋势检验

多组生存率比较时, 若分组变量是等级变量, 在 log-rank 检验组间生存率差别有统计学意义后, 还可以做趋势检验 (trend test), 分析危险率是否有随分组等级变化而变化的趋势。具体步骤如下。

(1) 按某种因素影响大小将患者分组, 可采用临床上该因素的自然分组。

(2) 计算每组的实际死亡数 A 与期望死亡数 T 。

(3) 进行趋势检验, 检验统计量。

$$\chi^2 = \frac{[\hat{S}(A - T)]^2}{\sum S^2 T - [(\sum ST)^2 / (\sum T)]} \quad (11-19)$$

服从自由度为 $k - 1$ 的 χ^2 分布。

11.3.3 Breslow 检验

Breslow 检验也称为 Breslow 广义 Wilcoxon 检验。

检验统计量 χ^2 :

$$\chi^2 = \frac{(\sum \omega_i d_{ki} - \sum \omega_i T_{ki})^2}{\sum \omega_i^2 V_{ki}} \quad (11-20)$$

其中, 权重 $\omega_i = n_i$ (即各期初例数合计值)。 n_i 随着生存时间的组增大逐渐减小, 所以 Breslow 检验给观察早期的差别更大的权重, 也就是对观察早期差别敏感, 而 log-rank 检验由于给观察后期的差别更大的权重, 即对于观察后期差别敏感。

11.4 Cox 比例风险模型

11.4.1 Cox 回归模型的构造

由于生存时间通常并不服从正态分布, 而且常会出现截尾数据, 不满足线性回归的模型要求, 所以采用的多因素分析方法是比例危险率回归模型分析, 简称 Cox 回归, 属于半参数模型。Cox 模型以生存结局和生存时间的危险率函数为因变量, 可同时分析众多因素对生存期的影响, 分析带有删失生存时间的资料, 且不要求资料服从特定的分布类型。

Cox 模型必须满足比例风险假定, 即任何两个个体的风险函数之比 (即风险比) 保持为一个与时间无关的恒定的比例, 模型中的协变量效应不随时间改变而改变。当不满足比例风险假定时, 可以采用分层比例风险模型。

Cox 模型不是直接考虑生存函数 $S(t, X)$ 与协变量的关系, 而是用风险率函数 $h(t, X)$ 作为因变量, 并假定

$$h(t, X) = h_0(t) e^{\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m} \quad (11-21)$$

由于公式是具有协变量 X 的个体在时刻 t 时的风险函数, 又称为瞬时死亡率。 t 表示生存时间, $X = (X_1, X_2, \cdots, X_m)'$ 表示与生存时间可能有关的协变量或交互项。 $h_0(t)$ 是所有危险因素为 0 时的基础风险率, 是非参数部分, 与 $h(t, X)$ 成比例。 $\beta = (\beta_1, \beta_2, \cdots, \beta_m)'$ 为 Cox 模型的回归系数, 需根据实际的观察数据估计。

回归系数 $\beta_i > 0$ 时, 协变量取值越大, 风险函数值越大, 表示患者死亡的风险越大, 反之越小, 当 $\beta_i = 0$ 时, 表示协变量对风险函数没有影响。

β_i 有明确的流行病学意义是在其他协变量不变的情况下, 协变量 X_i 每改变一个测定单位所引起的相对危险度的自然对数的改变量。相对危险度是风险函数 (也称危险度) 之比, 即

$$h(t, X_i)/h(t, X_j) = \exp[\beta'(X_i - X_j)] \quad (11-22)$$

按式 (11-22), 当协变量 X_i 取值为 0, 1 时, $RR = \exp(\beta_i)$; 当协变量 X_i 取值为连续型变量时, 用 X_i 和 X_i^* 分别表示在不同情况下的取值, $RR = e^{[\beta(x_i - x_i^*)]}$ 。

从 Cox 模型可以看出, 病人的风险率与该病人所有的危险因素及各因素对应的回归系数有关。回归系数反映因素对生存时间影响的强度, 标准回归系数可以比较不同因素间对生存时间的影响程度, 并且标准回归系数较大的因素对生存时间的影响也较大。个体预后指数 (personal prognosis Index) 为 $PI = \beta'_1 X'_1 + \beta'_2 X'_2 + \cdots + \beta'_m X'_m$, $PI = 0$ 表示该病人达到平均水平; $PI > 0$ 表示该病人的危险度大于平均水平, $PI < 0$ 表示该病人的危险度小于平均水平, 也可以表示为 $PI = \beta'_1 X'_1 + \beta'_2 X'_2 + \cdots + \beta'_m X'_m$, 其中那个 β' 为标准化回归系数, X' 为变量的标准化值。

可见, 对两个个体相对风险的评估不需要考虑基础风险函数, 所以 Cox 回归模型属于半参数回归模型; 风险函数的比值不会随时间的改变而改变, 所以又称为 Cox 比例风险模型, 或简称比例风险模型; 建立模型后, 可以利用公式 $HR(t | x = 1, x = 0) = e^{\beta(1-0)} = e^\beta$ 评估相对风险。

对回归模型的假设检验通常采用得分检验(score test)、Wald 检验($\chi^2 = \left(\frac{b_1}{S_{b_1}}\right)^2$)和似然比检验(maximum likelihood ratio test)($\chi^2 = 2[\ln L(p+1) - \ln L(p)]$), 这些检验统计量全部是 χ^2 检验, 自由度为模型中待检验的协变量个数。影响生存时间的因素称为协变量, 当协变量较多时, 可以先使用 χ^2 检验、log-rank 检验等对协变量进行筛选, 对有意义的协变量再进行 Cox 模型分析。模型中回归系数可借助偏似然函数(partial likelihood function)

$$L(\beta) = \prod_{i=1}^n \left[\frac{\exp(\beta_1 X_{i1} + \cdots + \beta_m X_{im})}{\sum_{S \in R(t_i)} \exp(\beta_1 X_{s1} + \cdots + \beta_m X_{sm})} \right] \quad (11-23)$$

进行估计。多因素分析时协变量的筛选策略与其他回归模型类似, 通常可采用逐步法。对于 Cox 模型一般情况下初筛的水准确定在 0.1 或 0.15, 设计严格的研究可确定为 0.05。

生存率与基础生存率相关, 故先采用非参数法或 Breslow 法估计出基础生存率, 再结合协变量的回归系数估计生存率

$$S(t, X) = [S_0(t)]^{\exp(\beta'X)} \quad (11-24)$$

在进行多因素 Cox 回归分析时, 其自变量的赋值方法与多重线性回归、多因素 logistic 回归分析方法相似, 对多值名义分类变量或多值有序分类变量进行赋值时, 可以采用哑变量赋值方法。具体赋值方法在前面已介绍过, 在此不再详述。

11.4.2 Cox 回归模型的拟合步骤

模型拟合及分析步骤如下。

- (1) 将收集到的资料进行整理成包括影响因素、随访时间以及结局的生存数据表。
- (2) 确定回归变量及变量的取值: 当遇到分类变量时, 常会进行哑变量的转化。
- (3) Cox 回归系数的估计: 在 SAS 软件中, 这个过程通过 PHREG 过程实现。如果生存时间有重复, 常使用 Breslow 和 Efron 提出的方法估计回归系数。

11.4.3 偏似然函数及参数估计的计算方法

为了避免对基础风险函数做确切估计, 可以使用偏似然函数解决 Cox 回归模型的参数估计。偏似然函数构造的思想是: 对每一个死亡点 t_i , 其风险集合 R_i 中的每一个体经历这一时间点时均有死亡危险, 而实际是第 i 个个体死亡的风险构造条件死亡概率, 再根据概率乘法规则, 各时间点条件死亡概率均发生的概率为各时间点条件死亡概率的乘积构造出的偏似然函数。偏似然函数表达式:

$$L(\beta) = \prod_{i=1}^n \left[\frac{e^{X_i'\beta}}{\sum_{j \in R_i} e^{X_j'\beta}} \right]^{c_i} = \prod_{i=1}^n \frac{e^{X_i'\beta}}{\sum_{j \in R_i} e^{X_j'\beta}} \quad (11-25)$$

式中, R_i 表示在时间 t_i 上的风险集合, 也就是将 n 名患者的生存时间按由小到大的顺序排列得到以后有序时间序列, 凡生存时间等于及大于 t_i 的所有患者组成的在时间 t_i 上的风险集合。 X_i 表示恰在 t_i 死亡的那个患者的回归向量。

用 Newton-Raphson 迭代法求解最大偏似然估计, 有信息矩阵:

$$I(\beta) = - \frac{\partial^2 \ln L(\beta)}{\partial \beta_r \partial \beta_s} \quad r, s = 1, 2, \cdots, m \quad (11-26)$$

回归系数的方差估计:

$$\hat{V}(\hat{\beta}) = I(\hat{\beta})^{-1} \quad (11-27)$$

回归系数的标准误估计值:

$$\hat{S}(\hat{\beta}) = [I(\beta)^{-1}]^{1/2}$$

(11-28)

当参数估计针对的是生存时间有重复 (ties) 时, 需要对上述的偏似然函数进行修正, 可以使用精确偏似然函数, 也可以使用 Breslow 和 Efron 的方法, 公式分别为:

$$L(\beta) = \prod_{i=1}^k \frac{e^{Z_i'\beta}}{(\sum_{j \in R_i} e^{X_j'\beta})^{d_i}}$$

(11-29)

$$L(\beta) = \prod_{i=1}^k \frac{e^{Z_i'\beta}}{\prod_{j=1}^{d_i} \left(\sum_{j \in R_i} e^{X_j'\beta} - \frac{f-1}{d_i} \sum_{j \in R_i} e^{X_j'\beta} \right)}$$

(11-30)

11.4.4 回归系数的统计学意义检验

对 Cox 回归模型做统计学意义检验的方法有偏似然比检验 (partial likelihood test)、Wald 检验和 Score 检验。

偏似然比检验的统计量:

$$\chi^2_{LR} = 2[\ln L(\hat{\beta}) - \ln L(\hat{\beta}_{-r})]$$

(11-31)

其中, $\ln L(\hat{\beta})$ 表示模型中包括全部回归变量时的对数偏似然函数, $\ln L(\hat{\beta}_{-r})$ 表示模型包括 $m-1$ 个回归变量但不包括第 r 个回归变量时的对数似然函数。在无效假设 $\beta_r = 0$ 成立的前提下, 服从自由度为 1 的 χ^2 分布。

Wald 检验统计量:

$$z = \frac{\hat{\beta}_r}{\hat{S}(\hat{\beta}_r)}$$

(11-32)

大样本, z 服从标准正态分布。 z 的平方服从自由度为 1 的 χ^2 分布, 所以, 又可以表示成:

$$\chi^2_w = \left[\frac{\hat{\beta}_r}{\hat{S}(\hat{\beta}_r)} \right]^2$$

(11-33)

实际应用中, 三种检验方法得到的统计学检验结果应该是一致的, Wald 检验计算比较简便, 当样本有限时, 偏似然比检验结果比较稳定, 作为首选。

11.5 三种常见回归模型的选择

多重线性回归、logistic 回归以及 Cox 回归模型是我们在解决实际问题中经常会使用到的方法, 归纳相同点和不同点如表 11-2 所示。

表 11-2 三种回归模型的不同点和相同点

多重线性回归		logistic 回归	Cox 回归
因变量	连续变量	分类变量	生存时间
因变量分布	正态分布	二项分布	无特定要求
删失	不允许	不允许	允许
模型结构	$Y = \beta_0 + \sum \beta_j X_j$	$\text{logit}(p) = \beta_0 + \sum \beta_j X_j$	$h(t) = h_0(t) \exp(\sum \beta_j X_j)$
参数估计	最小二乘法	最大似然法	最大似然法
参数检验	F 检验; t 检验	似然比检验; 计分检验; Wald 检验	似然比检验; 计分检验; Wald 检验

续表

	多重线性回归	logistic 回归	Cox 回归
	其他变量不变条件下, 变量 X_i	其他变量不变条件下, 变量 X_j	其他变量不变条件下, 变量 X_j
参数解释	每增加一个单位所引起的 Y 的平均改变量	每增加一个单位所引起的优势比 OR 的自然对数	每增加一个单位所引起的相对危险度 RR 的自然对数
预测指标	$\hat{Y}(-\infty \leq \hat{Y} \leq \infty)$	$\hat{p}(0 \leq \hat{p} \leq 1)$	$\hat{S}(t)(0 \leq \hat{S}(t) \leq 1)$
样本含量	至少 5 ~ 10 倍的自变量个数	至少 15 ~ 20 倍的自变量个数	至少 15 ~ 20 倍的自变量个数
变量筛选	前进法、后退法、逐步法	前进法、后退法、逐步法	前进法、后退法、逐步法

1. 自变量可以是连续变量、有序分类变量和无序分类变量, 无序分类变量带入回归方程需要进行哑变量化;
2. 自变量之间存在较强的相关关系时可能导致共线性现象;
3. 自变量之间可能存在交互作用;
4. 均可采用逐步回归筛选变量, 思路是先考虑单变量分析再进行多元回归分析;
5. 可以进行影响因素的分析、混杂因素的校正以及预测分析等。

11.6 案例分析

【例 11-1】 用某方案治疗大细胞肺癌与鳞状细胞肺癌, 随访观察获得每组个体的生存时间(天)和随访结果如表 11-3 (数字后的“+”表示删失, 其他为死亡), 试进行生存率的比较。

表 11-3 某方案治疗两种肺癌的生存时间和结果(天)

大细胞肺癌	11	32 +	42	42	71	115	124	174 +	186	278
鳞状细胞肺癌	60	102	165	189 +	200	254	281	295 +	458	

假设检验:
 H_0 : 样本所来自的总体生存分布相同。
 H_1 : 样本所来自的总体生存分布不相同。
SAS 程序:

```
data nkkysj11_1;
  do group=1 to 2;
    input n;
    do i=1 to n;
      input x censor @@ ;
      output;
    end;
  end;
cards;
10
11 0 32 1 42 0 42 0 71 0 115 0 124 0 174 1 186 0 278 0
9
60 0 102 0 165 0 189 1 200 0 254 0 281 0 295 1 458 0
;
run;
proc lifetest data =nkkysj11_1 method=pl plot = (s);
  time x*censor(1);
  strata group;
run;
```

程序说明:
数据输入时使用两个变量表示生存期, x 表示观测到的时间, censor 表示终点状态, 用 0 表示完全数据, 1 表示截尾数据。

Lifetest 过程求生存函数, method = pl 表示要求进行乘积极限法 (Product-Limit) 求函数。plot = (s)画出生存函数图。Time 语句指定生存期变量, 格式: 生存期变量名 * 表示终点状态的变量名(代表截尾的值)。Strata 语句要求对所指定的分组变量各组分别计算生存函数并做各组生存函数的统计学检验。如果没有设置 strata 语句, 意味着不对变量进行分组, 也就是对全部数据计算生存函数。

SAS 结果:

Stratum 1: group = 1					
Product-Limit Survival Estimates					
x	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.000	1.0000	0	0	0	10
11.000	0.9000	0.1000	0.0949	1	9
32.000 *	.	.	.	1	8
42.000	.	.	.	2	7
42.000	0.6750	0.3250	0.1551	3	6
71.000	0.5625	0.4375	0.1651	4	5
115.000	0.4500	0.5500	0.1660	5	4
124.000	0.3375	0.6625	0.1581	6	3
174.000 *	.	.	.	6	2
186.000	0.1688	0.8313	0.1431	7	1
278.000	0	1.0000	0	8	0

这是第一组(大细胞肺癌)的生存函数计算结果。以上部分为针对大细胞肺癌 (group = 1) 的生存函数估计值列表以及有关的生存时间统计量。列表从左到右依次为生存时间(X)、生存率(Survival)、死亡概率(Failure)、生存率标准误(Survival Standard Error)、死亡人数(Number Failed)、剩余人数(Number Left), 生存时间中标有星号“*”的为删失值。主要看生存时间(x), 生存率(Survival)和生存率的标准误(Survival Standard Error)。

以下部分列出了第一组(大细胞肺癌)生存时间变量的汇总统计量, 还有四分位数及 95% 可信区间等。最后给出了如下生存时间的平均数及其标准误。

Summary Statistics for Time Variable x			
Quartile Estimates			
Percent	Point Estimate	95% Confidence Interval	
		[Lower	Upper)
75	186.000	71.000	278.000
50	115.000	42.000	186.000
25	42.000	11.000	124.000

Mean	Standard Error
123.725	32.003

第二组(鳞状细胞肺癌)的生存函数计算结果如下。

Stratum 2: group = 2					
Product-Limit Survival Estimates					
x	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.000	1.0000	0	0	0	9
60.000	0.8889	0.1111	0.1048	1	8
102.000	0.7778	0.2222	0.1386	2	7
165.000	0.6667	0.3333	0.1571	3	6
189.000 *	.	.	.	3	5
200.000	0.5333	0.4667	0.1733	4	4
254.000	0.4000	0.6000	0.1738	5	3
281.000	0.2667	0.7333	0.1590	6	2
295.000 *	.	.	.	6	1
458.000	0	1.0000	0	7	0

第二组(鳞状细胞肺癌)生存函数的百分位数、均数及其标准误。

Summary Statistics for Time Variable x			
Quartile Estimates			
Percent	Point Estimate	95% Confidence Interval [Lower Upper)	
75	458.000	200.000	458.000
50	254.000	165.000	458.000
25	165.000	60.000	254.000

Mean	Standard Error
256.467	51.862

输出如下各组及总的例数、死亡数、截尾数和截尾百分比。

Summary of the Number of Censored and Uncensored Values					
Stratum	group	Total	Failed	Censored	Percent Censored
1	1	10	8	2	20.00
2	2	9	7	2	22.22
Total		19	15	4	21.05

下面是对两组生存曲线统一性的假设检验结果。

Testing Homogeneity of Survival Curves for x over Strata		
Rank Statistics		
group	Log-Rank	Wilcoxon
1	3.3724	42.000
2	-3.3724	-42.000

首先给出了两种秩检验统计量，包括如下 log-rank 秩检验和 Wilcoxon 秩检验。

Covariance Matrix for the Log-Rank Statistics		
group	1	2
1	2.79637	-2.79637
2	-2.79637	2.79637

Covariance Matrix for the Wilcoxon Statistics		
group	1	2
1	475.000	-475.000
2	-475.000	475.000

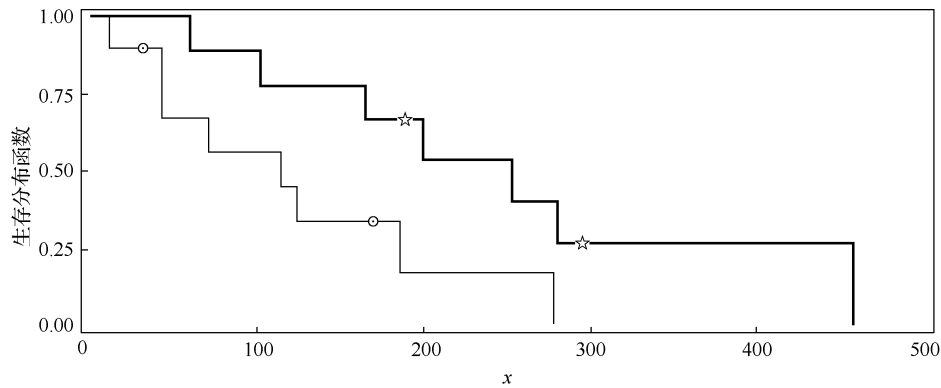
随后为两种秩检验的协方差矩阵。

Test of Equality over Strata			
Test	Chi-Square	DF	Pr >Chi-Square
Log-Rank	4.0670	1	0.0437
Wilcoxon	3.7137	1	0.0540
-2Log(LR)	2.1213	1	0.1453

输出两组生存函数统计学检验的结果。常使用的是 log-rank 检验，卡方值为 4.0670， $P =$

0.0437, 两组生存函数的差异有统计学意义。 $\log\text{-rank}\chi^2$ 统计量当生存时间符合 weibull 分布或属于比例风险模型时, 检验统计效率最高。 $\text{Wilcoxon}\chi^2$ 统计量当生存时间符合对数正态分布时, 检验统计效率较高。

生存率曲线图(如图 11-1 所示)直观地反映了两组患者生存率的差别, 可以看到某治疗方法对于鳞状细胞肺癌的生存时间明显优于大细胞肺癌的患者。



注: —— 代表第一组 ○ 代表第一组中的删失观测值, 代表第二组 ☆ 代表第二组中的删失观测值

图 11-1 生存率曲线图

【例 11-2】 探讨某恶性肿瘤的预后, 收集了 60 例病人的生存时间、结局及影响因素。影响因素包括病人的治疗方式、肿瘤的浸润程度、组织学类型、是否有淋巴转移及病人的性别、年龄, 生存时间以月计算, 变量的赋值和收集的原始资料如表 11-4 所示, 数据见表 11-5。进行半参数生存分析。

表 11-4 某恶性肿瘤的影响因素与赋值

变 量	因 素	赋 值 说 明
X_1	年龄	实际年龄(岁)
X_2	性别	男=1, 女=0
X_3	肿瘤组织学类型	高分化=1, 低分化=0
X_4	治疗方式	传统治疗方式=1, 新治疗方式=0
X_5	肿瘤是否转移	是=1, 否=0
X_6	肿瘤浸润程度	突破浆膜层=1, 未突破浆膜层=0
t	生存时间	实际生存时间(月)
Y	患者结局	死亡=0, 删失=1

表 11-5 60 例某恶性肿瘤患者的生存时间(月)及影响因素

序号	X_1	X_2	X_3	X_4	X_5	X_6	t	Y	序号	X_1	X_2	X_3	X_4	X_5	X_6	t	Y
1	54	0	0	0	1	0	52	1	31	43	1	0	0	1	0	120	1
2	57	0	1	1	0	0	51	1	32	44	1	0	0	1	2	120	1
.....																	
29	32	1	1	0	0	0	120	1	59	74	0	0	1	1	1	7	0
30	45	0	1	1	0	1	120	1	60	61	0	1	1	1	0	40	0

采用逐步回归的方法进行 Cox 模型分析, 进入和剔除的水准分别定为 0.05 和 0.05。通过 SAS 程序给出筛选之后的最佳模型包含 X_4 和 X_5 , 默认情况下包含三种检验方法(似然比法、得分法和 Wald 法)对应的 P 值均表明在 0.05 的检验水准上均得出配合的 Cox 模型具有统计学意义。

SAS 程序如下：

```
data nkkysj11_2;
input n x1 x2 x3 x4 x5 x6 t y@@ ;
cards;
1 54 0 0 0 1 0 52 1 31 43 1 0 0 1 0 120 1
2 57 0 1 1 0 0 51 1 32 44 1 0 0 1 2 120 1
3 58 0 0 1 1 1 35 0 33 62 0 0 1 1 2 120 1
4 43 1 1 0 1 0 103 1 34 40 1 1 0 0 1 40 0
5 48 0 1 1 0 2 7 0 35 50 1 0 1 1 0 26 0
6 40 0 1 1 0 2 60 1 36 33 1 1 1 0 2 120 1
7 44 0 1 1 0 2 58 1 37 57 1 1 0 0 0 120 1
8 36 0 0 1 1 1 29 0 38 48 1 0 1 1 2 120 1
9 39 1 1 0 0 1 70 1 39 28 0 0 1 1 2 3 0
10 42 0 1 1 0 1 67 1 40 54 1 0 0 1 0 120 0
11 42 0 1 1 0 0 66 1 41 35 0 1 1 1 1 7 0
12 42 1 0 0 1 2 87 1 42 47 0 0 1 1 2 18 0
13 51 1 1 0 0 0 85 1 43 49 1 0 0 1 0 120 1
14 55 0 1 1 0 1 82 1 44 43 0 1 1 0 0 120 1
15 49 1 1 0 0 1 76 1 45 48 1 1 1 0 2 15 0
16 52 1 1 0 0 1 74 1 46 44 0 0 1 1 2 4 0
17 48 1 1 0 0 2 63 1 47 60 1 1 0 0 2 120 1
18 54 1 0 0 1 1 101 1 48 40 0 0 1 1 2 16 0
19 38 0 1 1 0 0 100 1 49 32 0 1 1 0 1 24 0
20 40 1 1 0 0 1 66 0 50 44 0 0 1 1 1 19 0
21 38 0 0 1 1 2 93 1 51 48 1 0 1 1 0 120 1
22 19 0 0 1 1 2 24 0 52 72 0 1 1 1 2 24 0
23 67 1 0 0 1 0 93 1 53 42 0 0 1 1 0 2 0
24 37 0 0 0 1 0 90 1 54 63 1 0 0 1 0 120 1
25 43 1 0 1 1 2 15 0 55 55 0 1 0 0 2 120 1
26 49 0 0 1 1 2 3 0 56 39 0 0 1 1 2 50 0
27 50 1 1 0 1 1 87 1 57 44 0 0 1 1 0 120 1
28 53 1 1 0 0 2 120 1 58 42 1 1 0 0 2 120 1
29 32 1 1 0 0 0 120 1 59 74 0 0 1 1 1 7 0
30 45 0 1 1 0 1 120 1 60 61 0 1 1 1 0 40 0
;
run;
ods html;
proc phreg;
model t* y(1) =x1 -x6/selection=stepwise sle=0.05 sls=0.05 risklimits;
run;
ods html close;
```

SAS 程序说明：

数据集中共有 9 个变量，第 1 个是编号，第 2~6 个是影响生存时间的因素变量，t 是生存事件，y 是截尾变量，并且 1 为删失数据，0 表示完整数据。Cox 比例风险模型借助 Proc phreg 进行，model 语句中 = 前面定义生存时间变量、截尾变量及其截尾数据所对应的变量值，后面是影响因素变量，选项 selection 选择变量筛选的方式，SLS 和 SLE 分别表示入选标准和剔除标准。

SAS 主要运行结果：

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
60	23	37	61.67

Step 1. Variable x4 is entered. The model contains the following explanatory variables:

Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	177.183	161.941
AIC	177.183	163.941
SBC	177.183	165.077

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	15.2418	1	<.0001
Score	13.7383	1	0.0002
Wald	10.1295	1	0.0015

Step 2. Variable x5 is entered. The model contains the following explanatory variables:

Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	177.183	155.312
AIC	177.183	159.312
SBC	177.183	161.583

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	21.8707	2	<.0001
Score	19.7343	2	<.0001
Wald	15.4884	2	0.0004

这一部分给出生存分析信息，以及利用逐步回归的方法进行变量筛选，一共进行了两步：第一步筛选出 x_4 ，第二步筛选出 x_5 。Phreg 过程对数据进行 Cox 回归。Model 语句定义 Cox 回归的因变

量和协变量。等号左侧是生存期，格式：生存期变量名 * 表示终点状态的变量名（代表截尾的值）。等号右侧为各协变量名。斜杠后面为各种选择项。“selection = stepwise”表示要进行逐步 Cox 回归。“sle = ”和“sls = ”分别指定选入和剔除变量的统计学检验水平。“risklimits”要求输出风险比 HR 的 95% 置信区间。

SAS 给出了对每一步的模型拟合统计量（-2LOGL、AIC 和 SBC）及模型检验结果（似然比检验、Score 检验和 Wald 检验），之后是模型的最大似然估计，包括参数估计（Parameter Estimate）、Wald χ^2 （Chi-Square）、P 值（Pr > ChiSq），风险比 HR 及 HR95% 置信区间（Hazard Ration Confidence Limits）。

Analysis of Maximum Likelihood Estimates								
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
x4	1	1.94640	0.62374	9.7378	0.0018	7.003	2.062	23.781
x5	1	1.19587	0.50853	5.5301	0.0187	3.306	1.220	8.958

Summary of Stepwise Selection						
Step	Variable		Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
	Entered	Removed				
1	x4		1	13.7383	.	0.0002
2	x5		2	6.1925	.	0.0128

这里 x_4 和 x_5 的回归系数分别为 1.94640 和 1.19587，相对危险度为 7.003 和 3.306。说明传统方式与新方法相比危险性较大，并且存在淋巴转移的病人危险性较大。Cox 回归分析的结果，得到风险函数表达式为：

$$h(t) = h_0(t) \exp(1.94640x_4 + 1.19587x_5)$$

此式的右边指数部分取值越大，风险函数越大，预后越差，故称为预后指数。

11.7 生存分析案例辨析

【例 11-3】 原文题目：肝硬化食管静脉曲张破裂出血患者预后 Cox 回归分析。目的：探讨影响肝硬化食管静脉曲张破裂出血患者预后的危险因素。统计学方法：建立肝硬化食管静脉曲张数据库，计量资料采用 $\bar{x} \pm s$ 表示，应用 SPSS 17.0 软件进行统计分析，采用单因素 Cox 回归分析筛选影响预后的危险因素，单因素分析有意义者（ $P < 0.1$ ）纳入多因素 Cox 回归分析（前进法）。

【实例辨析】 原文作者在采用 Cox 回归比例风险模型做多因素分析之前，先将全部的自变量逐个地分别进行单因素分析，挑选出 $P < 0.1$ 的自变量，然后，将这些筛选出的自变量选入多因素 Cox 比例风险模型中。其实，这种筛选自变量的策略是错误的。因为自变量与自变量之间可能存在交互作用。因此，单变量分析的结果在很多情况下是不可靠的，因而其结果只是起参考作用，并不能反映其真实效应。

筛选自变量的正确策略是：当自变量不是特别多时，不需先进行单因素筛选，应尽可能使全部

自变量都有机会参与多因素模型的变量筛选过程；当自变量特别多且样本含量又不太大时(一般要求样本含量为自变量个数的 5 ~ 10 倍以上，回归分析的结果才较稳定)可以将单个自变量筛选中“ $P > 0.5$ ”的那些自变量暂时不参与自变量的筛选过程，必要时可从其中随机抽取几个与拟参与自变量筛选的那些自变量一并考察。筛选自变量的方法很多(SAS 在进行多重线性回归分析的过程中编入了 8 种筛选自变量的方法)，最好用 3 ~ 5 种方法处理同一份资料。若计算结果比较稳定，则可做出相应的统计学和专业结论。否则，宜考虑采用最优回归子集法筛选变量，并结合最优回归模型的评价标准和专业知识，确定 1 ~ 2 个最为理想的回归模型。

(孔雨佳 王素珍 李炳海)

第12章 判别分析

判别分析是另一种常用的多元统计分析方法，用于解决医学研究中需要判别分类的问题。例如，临床上需要根据患者的各项症状、实验室检查、病理学检查、体征及医学影像学资料诊断患者所患疾病类型。

12.1 判别分析的基本思想

判别分析是依据已知类别的样本所提供的信息，总结出分类的规律性，建立判别函数，并利用这个函数来判别新的样本所属类型，是判别个体所属群体的一种多元统计学方法。

根据以上给出的判别分析的定义，我们可知判别分析是已知研究对象的若干分类，并已经取得每一个类别的一批样本数据，在此基础上探索出分类的规律性，建立判别公式，然后对未知类别的样品进行判别。例如，我们有了患慢性胃溃疡病人和健康人的一些化验指标，就可以发现这些化验指标在这两组人中的区别，将这种区别归纳为一个判别函数，理论上我们就可以利用这个公式来判别某人是否患慢性胃溃疡。前提条件是所检测的那些“化验指标”确实具有明显区分人体胃肠道内是否患有溃疡的能力。

判别分析内容很丰富，方法很多。判别分析按判别的组数来区分，有两组判别分析和多组判别分析；按区分不同总体所用的数学模型来分，有线性判别和非线性判别；按判别时处理变量所用的方法不同来分，有逐步判别和序贯判别等。判别分析可以从不同角度找出解决问题的切入点，因此有不同的判别准则，如马氏距离最小准则、Fisher 准则、平均损失最小准则、最小平方准则、最大似然准则、最大概率准则等，按判别准则的不同又提出多种判别方法。常用的判别方法有距离判别法、Fisher 判别法、Bayes 判别法、逐步判别法和 logistic 判别法。

12.2 判别分析的步骤

(1) 判别分析的前提条件。

- ①因变量的取值是独立的，并事先已知。
- ②自变量服从多元正态分布。
- ③所有自变量在各组间方差协方差矩阵相等。
- ④自变量间不存在多重共线性。

(2) 明确已知分类的样本及所使用指标。

(3) 根据样本数据构建判别函数。构建判别函数的方法很多，最常用的判别方法包括最大似然法、距离判别法、Fisher 判别法和 Bayes 判别法等。

(4) 验证所构建的判别函数判别的结果与实际类别是否一致。分为组内回顾性验证和组外前瞻性验证。验证指标包括特异度、灵敏度、假阳性率、假阴性率和总符合率。如果验证结果的诊断符合率达到临床要求，则说明该判别函数具有实际应用价值，可用于未知类别样品的判别。

判别函数的一般形式为：

$$Y = a_1X_1 + a_2X_2 + a_3X_3 + \cdots + a_nX_n$$

(12-1)

其中, Y 是判别分数, X_1, X_2, \cdots, X_n 是反映研究对象特征的自变量, a_1, a_2, \cdots, a_n 为各变量的系数, 称为判别系数。可以看出, 这里我们假定的判别函数为线性判别函数。

(5) 样品判别分类。判别分析的目的就是对未知分类样品进行归类, 因此, 根据经过验证的有效的判别函数进行样品的分类判别是判别分析的最后一步。

12.3 实例分析

【例 12-1】 为研究舒张压 X_1 (kPa) 和血浆胆固醇 X_2 (mmol/L) 对冠心病的作用, 在某工厂测定了 50~59 岁女工冠心病患者 15 例和健康人 16 例的舒张压和血浆胆固醇, 结果见表 12-1。试做判别分析。

表 12-1 健康人和冠心病患者舒张压与血浆胆固醇测定值

NO	g	X_1	X_2	NO	g	X_1	X_2
健康人				冠心病病人			
1	1	10.66	2.07	1	2	9.86	5.18
2	1	12.53	4.45	2	2	13.33	3.73
3	1	13.33	3.06	3	2	14.66	3.89
4	1	9.33	3.94	4	2	9.33	7.10
5	1	10.66	4.45	5	2	12.80	5.49
6	1	10.66	4.92	6	2	10.66	4.09
7	1	9.33	3.68	7	2	10.66	4.45
8	1	10.66	2.77	8	2	13.33	3.63
9	1	10.66	3.21	9	2	13.33	5.96
10	1	10.66	5.02	10	2	12.00	5.70
11	1	10.40	3.94	11	2	14.66	6.19
12	1	9.33	4.92	12	2	13.33	4.01
13	1	10.66	2.69	13	2	12.80	4.01
14	1	10.66	2.43	14	2	13.33	3.63
15	1	11.20	3.42	15	2	12.33	5.96
16	1	9.33	3.63				

判别分析的 SAS 程序为:

```
data nkkysj12_1;
input g X1 X2 @@ ;
cards;
1 10.66 2.07 1 12.53 4.45 1 13.33 3.06 1 9.33 3.94 1 10.66 4.45 1 10.66 4.92 1 9.33
3.68 1 10.66 2.77 1 10.66 3.21 1 10.66 5.02 1 10.40 3.94 1 9.33 4.92 1 10.66 2.69 1
10.66 2.43 1 11.20 3.42 1 9.33 3.63 2 9.86 5.18 2 13.33 3.73 2 14.66 3.89 2 9.33 7.
10 2 12.80 5.49 2 10.66 4.09 2 10.66 4.45 2 13.33 3.63 2 13.33 5.96 2 12.00 5.70 2
14.66 6.19 2 13.33 4.01 2 12.80 4.01 2 13.33 3.63 2 12.33 5.96
;
run;
proc discrim method=normal manova listerr pool=yes crosslisterr;
class g;
var X1 X2;
run;
```

程序说明:

调用 discrim 过程步进行判别分析。选项 method = normal，表示资料服从正态分布时所选方法，否则选 method = npar；manoval 选项表示要求进行两类多变量均值向量多元方差分析；listerr 选项要求输出训练样品回代符合率和个体样本的错误分类率；pool = yes 选项要求直接合并方差和协方差矩阵（不检验是否满足方差齐性），crosslisterr 选项要求输出刀切法测试得到的个体样品的错误分类信息和判别公式外推正确率。

主要输出结果：

（第一部分）

Multivariate Statistics and Exact F Statistics					
	S = 1	M = 0	N = 13		
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.47309308	15.59	2	28	<.0001
Pillai's Trace	0.52690692	15.59	2	28	<.0001
Hotelling-Lawley Trace	1.11374897	15.59	2	28	<.0001
Roy's Greatest Root	1.11374897	15.59	2	28	<.0001

（第二部分）

Linear Discriminant Function for g		
Variable	1	2
Constant	-46.56378	-67.82708
X ₁	6.75083	8.00992
X ₂	5.83609	7.41829

（第三部分）

Resubstitution Results using Linear Discriminant Function					
Posterior Probability of Membership in g					
Obs	From g	Classified into g	1	2	
2	1	2 *	0.1745	0.8255	
3	1	2 *	0.4105	0.5895	
10	1	2 *	0.4747	0.5253	
17	2	1 *	0.6577	0.3423	
22	2	1 *	0.7974	0.2026	
23	2	1 *	0.6901	0.3099	

* Misclassified observation

Number of Observations and Percent Classified into g			
From g	1	2	Total
1	13	3	16
	81.25	18.75	100.00
2	3	12	15
	20.00	80.00	100.00
Total	16	15	31
	51.61	48.39	100.00
Priors	0.5	0.5	

（第四部分）

Cross-validation Summary using Linear Discriminant Function					
Posterior Probability of Membership in g					
Obs	From g	Classified into g	1	2	
2	1	2 *	0.1233	0.8767	
3	1	2 *	0.2861	0.7139	
6	1	2 *	0.4735	0.5265	
10	1	2 *	0.4258	0.5742	
17	2	1 *	0.7612	0.2388	
22	2	1 *	0.8487	0.1513	
23	2	1 *	0.7383	0.2617	
* Misclassified observation					
Number of Observations and Percent Classified into g					
From g	1	2	Total		
1	12	4	16		
	75.00	25.00	100.00		
2	3	12	15		
	20.00	80.00	100.00		
Total	15	16	31		
	48.39	51.61	100.00		
Priors	0.5	0.5			

结果解释：

第一部分通过 4 个不同的统计量计算两类多指标均值向量之间的差异，发现差异有统计学意义。表示冠心病患者与健康人在多指标(舒张压和血浆胆固醇)均值向量间有统计学差异。第二部分给出判别函数的常数和系数，将两组的常数和系数相减，得到 Fisher 线性判别函数，如第二组减去第一组得：

$$Y = -21.2633 + 1.25909X_1 + 1.5822X_2$$

将某样品的各指标值 X_1 、 X_2 代入判别函数，若计算出的 $Y < 0$ ，则将该样品判为第一类；反之，若计算出的 $Y > 0$ ，则将该样品判为第二类；当 $Y = 0$ 时，则无法判断。第三部分为个体样品的错误分类率和回代符合率。本例有 6 例错分为不同组。第一组健康人判对 13 例，判错 3 例，判错率为 18.75%，冠心病患者组判对 12 例，判错 3 例，判错率为 20%。第四部分为刀切法考核样本判错率。结果显示第一组判对 12 例，判错率为 25%，第二组判错 3 例，判错率为 20%。

(曹文君 王素珍 杜宗刚)

第 13 章 聚类分析

聚类分析同判别分析一样,也是研究事物分类问题的多元统计分析方法。与判别分析不同的是,判别分析是依据已知类别样本提供的信息,总结分类规律,进而判别新的样品的归类;而聚类分析则是在不明确应分多少类的情形下,利用数理统计学方法采用已收集到的资料探索事物的适当分类的方法。

13.1 聚类分析的基本思想

聚类分析没有“历史资料”作为分类依据,只能根据研究对象本身的特性进行分类,是一种探索性统计分析方法,依照分类的目的分为样品聚类(又称 Q-型聚类)和指标聚类(R-型聚类)。其中 Q-型聚类,即样品聚类,是将样品(病例)归类的方法,其目的是找出样品间的共性;R-型聚类,即指标聚类,是将指标(变量)归类的方法,其目的是将指标降维从而选择有代表性的指标。

聚类分析的基本思想是相近的样品(或变量)先聚成类,相远的后聚成类,过程一直进行下去,每个样品(或指标)总能聚到合适的类中。因此,聚类分析的关键是度量两个指标或样品的相近程度。目前用得最多的方法是计算指标或样品间的相似系数。常用的 R-型聚类的相似系数包括简单相关系数 r 和针对分类资料的列联系数 C 。设 X_1, X_2, \dots, X_m 表示 m 个变量,如果这 m 个变量均服从正态分布,则用两个变量 X_i 和 X_j 间的 Pearson 相关系数 r_{ij} 定义相似系数。

$$r_{ij} = \frac{\left| \sum (X_i - \bar{X}_i)(X_j - \bar{X}_j) \right|}{\sqrt{\sum (X_i - \bar{X}_i)^2 \sum (X_j - \bar{X}_j)^2}} \quad (13-1)$$

如果 n 个变量不服从正态分布,则用两个变量间的 Spearman 秩相关系数 r_s 定义相似系数。

当变量均为定性变量时,用列联系数 C 定义相似系数。

$$C = \frac{\sqrt{\chi^2}}{\sqrt{\chi^2 + n}} \quad (13-2)$$

式中, χ^2 是 X_i (R 个水平) 和 X_j (C 个水平) 构成的 $R \times C$ 列联表的 χ^2 值, n 是总例数。相似系数绝对值越大表明两个变量间相似程度越高。常用的 Q-型聚类的相似系数用距离表示,距离的定义包括欧氏距离、绝对距离、马氏距离等多种定义,其中欧氏距离应用最为广泛,其计算公式为:

$$d_{ij} = \sqrt{\sum (X_i - X_j)^2} \quad (13-3)$$

样品聚类分析内容非常丰富,有系统聚类法、分解聚类法、动态样品聚类法等。系统聚类法是首先把 N 个对象看成 N 个类,然后将性质最接近的两类合并为一个新类,得到 $N-1$ 个类,再从中找出最接近的两个类合并成 $N-2$ 个类,如此下去,直到将 N 个对象合并成一个类为止;分解聚类法与系统聚类相反,首先将 N 个对象看成一类,然后按某种原则将它分成 2 类,进而分成 3 类……如此下去,直到每类仅剩一个对象为止;动态聚类法是首先按照某种法则将研究对象分成若干个类,然后依据某优化准则逐步调整,直到前后两次调整结果一致为止,类似于迭代算法。本章主要介绍常用的样品系统聚类法和指标聚类法。

13.2 样品系统聚类分析步骤

- 对样品进行系统聚类的一般步骤如下。
- (1) 确定待分类的样品。
 - (2) 收集数据并对数据进行变换处理(如标准化或规格化)。
 - (3) 将每个样品作为一类,即 n 个样品一共有 n 类。
 - (4) 计算各类之间的相似系数(即距离),得到一个相似系数对称矩阵,将相似系数最近的两个类并成一个新类。
 - (5) 合并后,重新计算当前 $n-1$ 个类之间的相似系数,再合并相似系数最近的两类,直至所有样品归为一类。
 - (6) 绘制系统聚类谱系图,按不同的分类标准并结合专业解释,得出不同的分类结果。

13.3 指标分解聚类分析步骤

- 指标分解聚类分析步骤如下。
- (1) 确定待分类的指标。
 - (2) 收集数据并对数据进行变换处理(如标准化或规格化)。
 - (3) 将所有指标聚为一类,计算该类的方差在总方差中所占的比例以及第二成分特征根的大小,以确定是否继续分类。若该类所占比例较小或第二成分特征值较大,则需继续分类。
 - (4) 选择类分量所占百分比最小或类分量第二特征值最大的类继续分类。
 - (5) 把选中的类划分成两个新类。
 - (6) 通过反复迭代使变量重新归类,最终使各分类所解释的方差达到最大。

13.4 实例分析

13.4.1 问题与数据

【例 13-1】 某医院测量了 27 名疑似骨结核患者的年龄和血常规 5 项指标(见表 13-1),请分别对此 27 例样本和 6 个变量进行聚类分析。

表 13-1 27 例疑似骨结核患者的年龄和血常规 5 项指标

编号	年龄(岁) X_1	血沉(mm/h) X_2	嗜中性粒细胞百分比(%) X_3	白细胞数($\times 10^9$ 个/L) X_4	血红蛋白(g/L) X_5	白球蛋白比例(%) X_6
1	50	25	78.62	8.36	132.0	0.82
2	30	32	75.30	6.83	114.7	0.79
3	35	99	67.10	10.59	130.1	1.04
4	34	16	68.03	3.70	136.0	1.15
5	10	16	58.21	6.44	108.5	0.99
6	24	75	69.45	5.90	145.0	1.24
7	22	6	75.28	9.49	148.4	1.46
8	23	6	49.63	5.25	147.0	1.11
9	63	90	69.59	5.11	111.0	1.13
10	20	10	66.69	10.5	129.0	1.75

续表

编号	年龄(岁) X_1	血沉(mm/h) X_2	嗜中性粒细胞百分 比(%) X_3	白细胞数 ($\times 10^9$ 个/L) X_4	血红蛋白(g/L) X_5	白球蛋白比 例(%) X_6
11	25	16	62.72	6.27	98.4	1.11
12	48	59	78.93	9.76	93.2	0.81
13	67	11	67.92	2.87	130.9	1.55
14	67	50	80.73	6.83	70.2	1.13
15	36	55	77.95	7.30	129.8	1.28
16	46	50	71.06	5.46	118.6	1.14
17	39	30	68.07	6.62	150.8	1.61
18	62	40	50.74	8.98	112.0	0.91
19	31	70	64.50	6.40	122.0	1.09
20	37	30	70.67	6.22	109.3	0.84
21	32	122	78.54	11.45	105.9	0.89
22	27	18	58.65	5.82	153.6	1.46
23	25	13	56.56	5.75	139.7	1.35
24	34	108	79.30	11.35	139.6	0.93
25	63	95	71.68	8.76	113.4	1.40
26	35	32	67.35	6.33	122.8	1.30
27	59	44	80.99	8.11	109.6	0.75

13.4.2 对给定资料进行无序样品聚类分析

27 例样品聚类分析的 SAS 程序为：

```
data nkkysj13_1;
input X1 -X6 @@ ;
cards;
50 25 78.62 8.36 132.0 0.82 30 32 75.30 6.83 114.7 0.79 3599 67.10 10.59 130.1 1.
04 34 16 68.03 3.70 136.0 1.15 10 16 58.21 6.44 108.5 0.99 24 75 69.45 5.90 145.0
1.24 22 6 75.28 9.49 148.4 1.46 23 6 49.63 5.25 147.0 1.11 63 90 69.59 5.11 111.0
1.13 20 10 66.69 10.5 129.0 1.75 25 16 62.72 6.27 98.4 1.11 48 59 78.93 9.76 93.2
0.81 67 11 67.92 2.87 130.9 1.55 67 50 80.73 6.83 70.2 1.13 36 55 77.95 7.30 129.8
1.28 46 50 71.06 5.46 118.6 1.14 39 30 68.07 6.62 150.8 1.61 62 40 50.74 8.98 112.
0 0.91 31 70 64.50 6.40 122.0 1.09 37 30 70.67 6.22 109.3 0.84 32 122 78.54 11.45
105.9 0.89 27 18 58.65 5.82 153.6 1.46 25 13 56.56 5.75 139.7 1.35 34 108 79.30 11.
35 139.6 0.93 63 95 71.68 8.76 113.4 1.40 35 32 67.35 6.33 122.8 1.30 59 44 80.99
8.11 109.6 0.75
;
run;
proc cluster method=average rsquare ccc pseudo outtree =tree;
var X1 -X6;
proc tree data =tree horizontal;
run;
```

程序说明：

调用 cluster 过程进行样品聚类分析。选项 method = average，表示选用类平均方法(一种距离定义法)进行聚类，其他可选项包括 centriod、median、ward 等；rsquare 选项要求输出半偏 R^2 和 R^2 ；ccc,pseudo 选项是为了计算一些统计量用以判断全部样品究竟聚成几类合适。其中，ccc 要求输出聚类判别据(即判断资料聚为几类合适的一种统计量)的立方值及在一致无效假设下近似期望值 R^2 ,pseudo 要求输出两个统计量伪 F 值(PSF)和伪 t^2 值(PsT2)；ccc 和 psf 出现峰值时(常为局部峰

值)所对应的分类数较合适, pst2 出现峰值(常为局部峰值)的前一行所对应的分类数较合适。out-tree = tree 选项表示输出结果到 tree 这个数据集中,它可被 tree 过程调用以输出聚类结果的树状图。

Var 语句是指出数据集中用来进行聚类分析的变量名称。如果省略,则 SAS 会自动对其他语句中未使用过的所有数值变量进行聚类分析。

调用 tree 过程,表明要对 cluster 过程建立的输出结果数据集 tree 做聚类树状图; horizontal 选项要求将图形水平放置。

主要输出结果:

(第一部分)

Average Linkage Cluster Analysis Eigenvalues of the Covariance Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	1239.12790	824.14719	0.6580	0.6580
2	414.98070	245.96320	0.2204	0.8783
3	169.01751	112.14688	0.0897	0.9681
4	56.87063	53.68722	0.0302	0.9983
5	3.18340	3.13182	0.0017	1.0000
6	0.05158		0.0000	1.0000
Root-Mean-Square Total-Sample Standard Deviation = 17.71643				
Root-Mean-SquareDistance Between Observations = 61.37152				

(第二部分)

Cluster History										Norm T	
										RMS i	
NCL	---	Clusters Joined	---	Freq	SpRSq	RSq	ERSq	CCC	Ps F	PsT2	Dist e
26 OB9		OB25		2	0.0005	1.00	.	.	80.7	.	0.1135
25 OB2		OB20		2	0.0011	.998	.	.	53.4	.	0.1661
24 OB8		OB23		2	0.0016	.997	.	.	41.5	.	0.2026
23 CL25		OB26		3	0.0021	.995	.	.	34.8	1.9	0.2173
22 CL24		OB22		3	0.0030	.992	.	.	28.9	1.9	0.2607
21 OB15		OB16		2	0.0031	.989	.	.	26.4	.	0.2828
20 OB3		OB24		2	0.0033	.985	.	.	25.0	.	0.2922
19 OB4		OB10		2	0.0034	.982	.	.	24.4	.	0.2957
18 OB5		OB11		2	0.0035	.979	.	.	24.2	.	0.3037
17 CL19		OB7		3	0.0054	.973	.	.	22.7	1.6	0.3561
16 CL17		CL22		6	0.0097	.964	.	.	19.4	2.9	0.3758
15 CL21		OB19		3	0.0070	.957	.	.	18.9	2.3	0.3946
14 OB1		OB17		2	0.0063	.950	.	.	19.1	.	0.4039
13 OB12		OB27		2	0.0064	.944	.	.	19.6	.	0.4064
12 CL23		CL18		5	0.0188	.925	.	.	16.8	8.5	0.4905
11 CL14		CL16		8	0.0230	.902	.	.	14.8	4.7	0.538

10	OB6	CL15	4	0.0143	.888	.	.	15.0	2.8	0.5393
9	CL20	OB21	3	0.0165	.871	.	.	15.2	5.0	0.5865
8	CL13	OB14	3	0.0158	.856	.	.	16.1	2.5	0.5917
7	OB13	OB18	2	0.0159	.840	.	.	17.5	.	0.6428
6	CL11	CL12	13	0.0682	.771	.	.	14.2	9.7	0.6627
5	CL9	CL26	5	0.0333	.738	.804	-2.4	15.5	4.9	0.6711
4	CL6	CL7	15	0.0501	.688	.754	-1.7	16.9	4.0	0.7906
3	CL5	CL10	9	0.0837	.604	.677	-1.7	18.3	7.5	0.8414
2	CL3	CL8	12	0.1198	.485	.524	-.64	23.5	6.5	1.0108
1	CL4	CL2	27	0.4845	.000	.000	0.00	.	23.5	1.2072

(第三部分)(图 13-1)

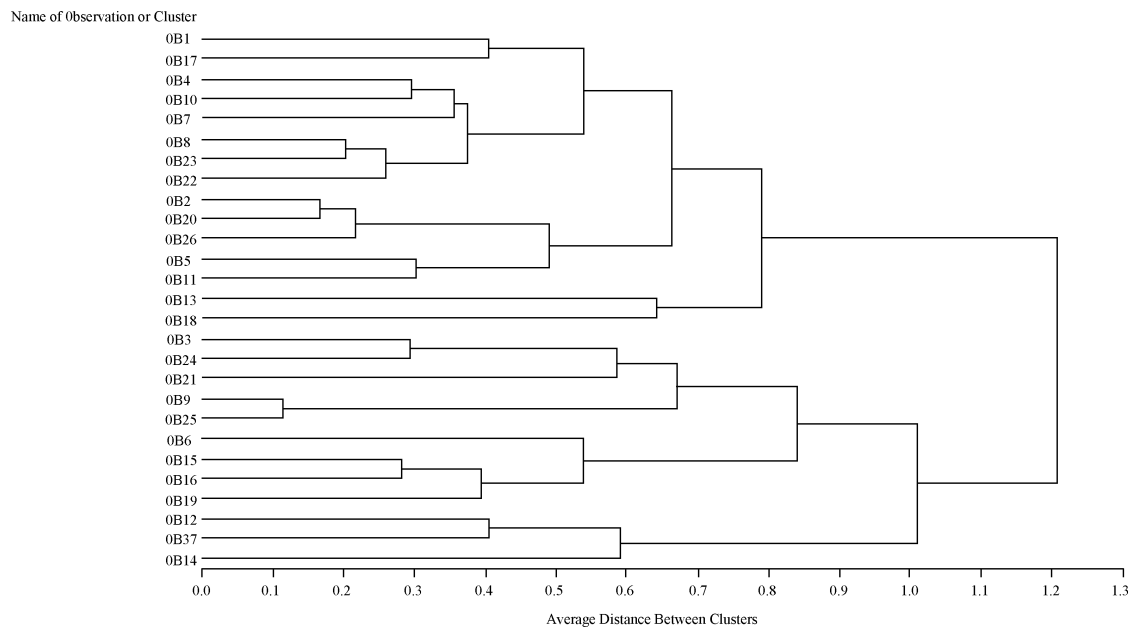


图 13-1 样品聚类的树状图

结果解释：

第一部分为 27 个样品协方差矩阵的特征值、两相邻特征值之差、各特征值占总方差的百分比和累计百分比。此结果有助于判断聚类数，注意到前 4 个特征值占总信息的 99.83%，提示样品可分为 4 类。结果第二部分为聚类过程，输出 R^2 、偏 R^2 、伪 F 和 t^2 等值。观察到 ccc、psf 和 pst2 均在 $NCL = 4$ 时达到局部的峰值，参考前面第 4 特征根的累计方差比达 99.83%，基本可以判定这 27 个样品分为 4 类合适。参考第三部分的样品聚类的树状图，可以把第 1、17、4、10、7、8、23、22、2、20、26、5、11、13、18 号患者归为一类，第 3、24、21、9、25 号患者归为一类，第 6、15、16、19 号患者归为一类，把第 12、27、14 号患者归为一类。

13.4.3 对给定资料进行变量聚类分析

6 个指标聚类分析的 SAS 程序为：


```
data nkkysj13_1;
input X1 -X6 @@ ;
cards;
50 25 78.62 8.36 132.0 0.82 30 32 75.30 6.83 114.7 0.79 35 99 67.10 10.59 130.1 1.
04 34 16 68.03 3.70 136.0 1.15 10 16 58.21 6.44 108.5 0.99 24 75 69.45 5.90 145.0
1.24 22 6 75.28 9.49 148.4 1.46 23 6 49.63 5.25 147.0 1.11 63 90 69.59 5.11 111.0
1.13 20 10 66.69 10.5 129.0 1.75 25 16 62.72 6.2798.4 1.11 48 59 78.939.76 93.2 0.
81 67 11 67.92 2.87 130.9 1.55 67 50 80.73 6.83 70.2 1.13 36 55 77.95 7.30 129.8 1.
28 46 50 71.06 5.46 118.6 1.14 39 30 68.07 6.62 150.8 1.61 62 40 50.74 8.98 112.0
0.91 31 70 64.50 6.40 122.0 1.09 37 30 70.67 6.22 109.3 0.84 32 122 78.54 11.45
105.9 0.89 27 18 58.65 5.82 153.6 1.46 25 13 56.56 5.75 139.7 1.35 34 108 79.30 11.
35 139.6 0.93 63 95 71.68 8.76 113.4 1.40 35 32 67.35 6.33 122.8 1.30 59 44 80.99
8.11 109.6 0.75
;
run;
proc varclus maxeigen=0.75 outtree=tree;
var X1 -X6; run;
proc tree data=tree horizontal;
run;
```

程序说明：

调用 SAS 中专门用于指标聚类的 Varclus 过程进行指标的聚类分析。选项 maxeigen = 0.75 表示每一类中第二特征值所允许的最大值，用来控制聚类的类别数，一般取 0.75，这个数取得越大，最后聚类分析得到的类别数就越小。

主要输出结果：
(第一部分)

Oblique Principal Component Cluster Analysis			
Observations	27	Proportion	0
Variables	6	Maxeigen	0.75

Clustering algorithm converged.

Cluster Summary for 1 Cluster					
Cluster	Members	Cluster Variation	Variation Explained	Proportion Explained	Second Eigenvalue
1	6	6	2.482032	0.4137	1.2410

Total variation explained = 2.482032 Proportion = 0.4137

Cluster 1 will be split because it has the largest second eigenvalue, 1.241028, which is greater than the MAXEIGEN = 0.75 value.

Clustering algorithm converged.

(第二部分)

Cluster Summary for 2 Clusters					
Cluster	Members	Cluster Variation	Variation Explained	Proportion Explained	Second Eigenvalue
1	3	3	1.725715	0.5752	0.8581
2	3	3	1.861101	0.6204	0.6439

Total variation explained = 3.586815 Proportion = 0.5978
R-squared with 2 Clusters

Cluster	Variable	Own Cluster	Next Closest	1 - R * * 2 Ratio
Cluster 1	X1	0.4438	0.0472	0.5838
	X5	0.7690	0.0823	0.2517
	X6	0.5129	0.1149	0.5503
Cluster 2	X2	0.6780	0.1275	0.3690
	X3	0.5531	0.1548	0.5288
	X4	0.6300	0.0167	0.3762

这里列出两类分别包含的指标，第一类包括 X_1 、 X_5 、 X_6 ，第二类包括 X_2 、 X_3 、 X_4 。第一类将被分解，因为该类的第二大特征值 0.858119 比最大特征界值 0.75 还要大。聚类算法收敛。

(第三部分)

Cluster Summary for 3 Clusters

Cluster	Members	Cluster Variation	Variation Explained	Proportion Explained	Second Eigenvalue
1	2	2	1.484229	0.7421	0.5158
2	3	3	1.861101	0.6204	0.6439
3	1	1	1	1.0000	

Total variation explained = 4.345329 Proportion = 0.7242
Oblique Principal Component Cluster Analysis
R - squared with 3 Clusters

Cluster	Variable	Own Cluster	Next Closest	1 - R * * 2 Ratio
Cluster 1	X5	0.7421	0.1890	0.3180
	X6	0.7421	0.1149	0.2914
Cluster 2	X2	0.6780	0.1039	0.3593
	X3	0.5531	0.1107	0.5026
	X4	0.6300	0.0422	0.3863
Cluster 3	X1	1.0000	0.1123	0.0000

Standardized Scoring Coefficients

Cluster	1	2	3
X1	0.00000	0.00000	1.00000
X2	0.00000	0.44243	0.00000
X3	0.00000	0.39959	0.00000
X4	0.00000	0.42650	0.00000
X5	0.58041	0.00000	0.00000
X6	0.58041	0.00000	0.00000

这是由变量预测类成分的标准回归系数，用 C_1 、 C_2 和 C_3 分别表示三个分类，则有：
 $C_1 = 0.58041X_5 + 0.58041X_6$ ， $C_2 = 0.44243X_2 + 0.39959X_3 + 0.42650X_4$ ， $C_3 = 1.00000X_1$

Cluster Structure			
Cluster	1	2	3
X1	-0.33518	0.21729	1.00000
X2	-0.32237	0.82341	0.27042
X3	-0.33278	0.74368	0.32645
X4	-0.20541	0.79375	-0.07691
X5	0.86146	-0.28680	-0.43469
X6	0.86146	-0.33897	-0.14280

聚类结构表示各变量由类成分构成的线性组合系数，如 $X_1 = -0.33518C_1 + 0.21729C_2 + 1.00000C_3$ 。

Inter-Cluster Correlations			
Cluster	1	2	3
1	1.00000	-0.36320	-0.33518
2	-0.36320	1.00000	0.21729
3	-0.33518	0.21729	1.00000

No cluster meets the criterion for splitting.

这个结果表示类与类间相关系数。

(第四部分)(图 13-2)

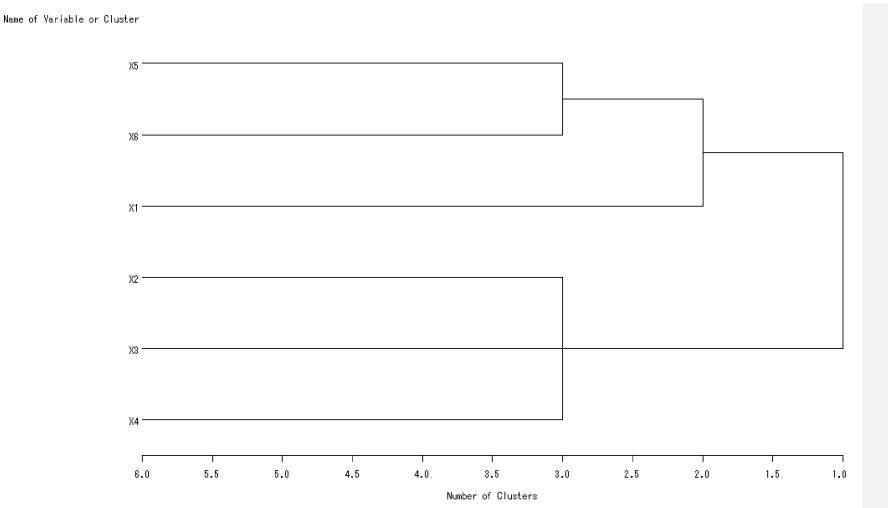


图 13-2 指标聚类的树状图

结果解释：第一部分将 6 个变量聚成一类，解释总方差的量为 2.482032，占总方差的 41.37%，需继续分解。第二部分是 将 6 个指标分成两类，分别包含 3 个指标，此时能解释总方差的量是 3.586815，占 59.78%，第一类中的第二特征值为 0.8581，大于 0.75，需进一步分类。第三部分是继续将第二部分中的第一类进行分解，共包含 3 类，此时能解释总方差的量是 4.345329，占 72.42%，各类中第二特征值均小于 0.75，分类截止。因此，本例 6 个变量聚成 3 类合适，其中 X_5 、 X_6 为一类， X_2 、 X_3 、 X_4 为一类， X_1 自成一类。

(曹文君 王素珍)

第 14 章 主成分分析

多指标是医学科学研究中经常遇到的问题之一，它既为研究者提供了丰富的信息，同时也增加了统计分析的困难。例如，在学生学习能力的评价中，收集到的数据包括每一个学生的数学、语文、英语等学科成绩，以及学生的智商等 10 多个指标。为了在尽量避免原指标信息损失的基础上，尽可能地减少分析指标，需采用特殊的统计分析方法，而主成分分析则是处理此类问题的有效方法之一。

主成分分析的基本思想是通过降维技术将多个变量综合成少数几个互不相关的主成分(本质上就是隐变量)的统计分析方法。这些主成分能够反映原始变量的绝大部分信息，它们通常是原始变量的某种组合。主成分分析的目的一般为变量的降维和间接实现样品聚类。

14.1 主成分分析的基本原理

设有 n 个观察对象，每个观察对象对应 p 个指标。数据资料可表示成以下矩阵形式：

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = (X_1, X_2, \cdots, X_p)$$

所谓主成分分析，就是寻找 $m(m < p)$ 个具有实际意义的相互独立的主成分 \mathbf{F} 来反映 $n \times p$ 个数据所蕴含信息的一种多元统计分析方法。一般来说， \mathbf{F} 能写成 \mathbf{X} 的线性组合，即 $\mathbf{F} = \mathbf{A}\mathbf{X}$ ，等价于式 (14-1)。

$$\begin{cases} \mathbf{F}_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p \\ \mathbf{F}_2 = a_{12}X_1 + a_{22}X_2 + \cdots + a_{p2}X_p \\ \vdots \\ \mathbf{F}_p = a_{1p}X_1 + a_{2p}X_2 + \cdots + a_{pp}X_p \end{cases} \quad (14-1)$$

其中，

$$X_i = [x_{1i}, x_{2i}, \cdots, x_{ni}]' \quad i = 1, 2, \cdots, p. \quad \mathbf{A} = \begin{Bmatrix} a_{11} & a_{21} & \cdots & a_{p1} \\ a_{12} & a_{22} & \cdots & a_{p2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{pp} \end{Bmatrix} = \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_p \end{pmatrix}$$

在寻找第一主成分 \mathbf{F}_1 时，我们希望 \mathbf{F}_1 能尽可能多地反映原来指标信息，一般用 \mathbf{F}_1 的方差来表达，方差越大，表示 \mathbf{F}_1 包含的信息越多。如果 $\mathbf{F}_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p = \mathbf{a}'_1\mathbf{X}$ 满足： $\mathbf{a}'_1\mathbf{a}_1 = 1$ ，且 $\text{Var}(\mathbf{F}_1) = \text{Max}\{\text{Var}(\mathbf{a}'_1\mathbf{X})\}$ ，则定义 \mathbf{F}_1 为原始指标 X_1, X_2, \cdots, X_p 的第一主成分。如果第一主成分不足以代表原来 p 个指标的信息，再考虑选取 \mathbf{F}_2 即选第二个线性组合，为了有效地反映原来信息，要求 \mathbf{F}_1 与 \mathbf{F}_2 相互独立，用数学语言表达为 $\text{Cov}(\mathbf{F}_1, \mathbf{F}_2) = 0$ ，称 \mathbf{F}_2 为第二主成分，依此类推可以构造出第三，第四， \cdots ，第 m 个主成分。

14.2 主成分的贡献率

数学上可以证明，主成分 F_1, F_2, \dots, F_p 的方差 $\lambda_1, \lambda_2, \dots, \lambda_p$ 是原始变量 X 的协方差矩阵(也就是变量标准化后的相关系数矩阵 R)的特征值，而 a'_j 就对应于 $\lambda_j (j = 1, 2, \dots, p)$ 的单位特征向量。第 i 分主成分的贡献率定义为第 i 个主成分解释的信息占全部 p 个变量总信息的比例，记为 η_i ，即

$$\eta_i = \lambda_i / \sum_{i=1}^p \lambda_i \tag{14-2}$$

显然，第一主成分 F_1 的贡献率最大，表明它提取原始变量 X_1, X_2, \dots, X_p 信息的能力最强，其次为 F_2 ，依此类推。一般，我们取 $m (m < p)$ 个主成分，则称前 m 个主成分的贡献率之和

$$P_m = \sum_{i=1}^m \lambda_i / \sum_{i=1}^p \lambda_i \tag{14-3}$$

为主成分 F_1, F_2, \dots, F_m 的累计贡献率。

如何确定主成分的个数是主成分分析的关键问题。目前，尚没有统一的标准来确定主成分的个数，在具体操作中，一般应遵循 $P_m > 75\%$ 或特征值大于 1，当然还要根据实际问题来确定，以保证主成分的可解释性。

14.3 主成分分析的步骤

(1) 按下式对原始指标数据进行标准化。

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, j = 1, 2, 3, \dots, p$$

X 标准化后的数据矩阵为：

$$X' = \begin{pmatrix} x'_{11} & x'_{12} & \cdots & x'_{1p} \\ x'_{21} & x'_{22} & \cdots & x'_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x'_{n1} & x'_{n2} & \cdots & x'_{np} \end{pmatrix}$$

(2) 计算 X' 的相关矩阵 R ，标准化后 X' 的相关矩阵即为协方差矩阵 $\text{Cov}(X')$ 。

$$R = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{np} \end{pmatrix}$$

(3) 计算相关矩阵 R 的特征值及特征向量，并将特征值按从大到小顺序进行排列，即

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$$

(4) 按照式(14-3)计算前 m 个主成分的累计贡献率，取前 m 个较大特征值对应的单位特征向量 $a_i = (a_{i1}, a_{i2}, \dots, a_{ip})'$ 组成变换矩阵，从而求得各主成分：

$$F_i = a'_i X = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p, i = 1, 2, \dots, m$$

(5) 主成分命名和解释。

说明：用统计软件实现主成分分析时，不必手工对各定量变量进行标准化，软件会自动实施。

14.4 实例辨析

【例 14-1】 采用主成分分析方法对糖尿病住院患者营养状况综合评价^[1]。目的：综合评价糖尿病住院患者营养状况，探讨科学的综合评价方法。方法：对 21 例肝肾功能正常无发热的糖尿病住院患者检测 8 个指标，如表 14-1 所示。

表 14-1 21 例肝肾功能正常无发热的糖尿病住院患者营养状况指标检测结果

编号 i	体重指数 x_1	臂肌围 x_2	血红蛋白 x_3	淋巴细胞 总数 x_4	总蛋白 x_5	白蛋白 x_6	前白蛋白 x_7	视黄醇结合 蛋白 x_8
1	20.18	18.77	132.34	3.23	64.92	34.89	342.53	1.19
2	23.29	24.30	138.57	2.29	51.24	47.57	367.00	59.30
3	22.44	20.18	133.99	1.45	53.45	39.96	72.27	60.60
4	22.14	20.77	126.70	1.86	72.76	37.64	283.20	61.97
5	16.81	23.95	145.45	2.77	75.84	49.97	339.38	39.65
6	17.13	20.31	132.99	2.70	64.12	45.26	328.60	48.78
7	22.29	20.06	132.88	0.39	72.83	41.20	398.70	32.94
8	29.24	22.41	136.95	3.19	84.82	46.24	170.90	31.15
9	22.76	18.53	112.37	1.62	70.02	50.40	301.97	22.80
10	22.29	20.79	92.72	2.28	76.12	39.62	343.92	39.65
11	22.21	23.24	144.92	1.88	57.03	31.05	355.15	43.95
12	19.29	17.60	149.52	0.91	59.76	37.50	448.10	54.87
13	28.50	20.15	138.62	2.58	59.80	45.92	208.91	64.65
14	20.32	19.00	116.02	1.59	67.23	38.10	369.67	69.77
15	21.83	20.07	124.53	1.28	71.71	31.61	338.32	44.80
16	17.06	18.17	170.35	0.49	57.66	38.18	320.88	33.03
17	17.99	23.45	120.88	2.50	85.35	42.40	459.46	65.16
18	27.03	23.93	123.96	1.64	62.85	33.37	304.17	68.40
19	26.04	17.98	149.27	1.33	57.51	37.91	273.26	27.99
20	25.50	20.72	147.79	1.31	73.29	43.69	176.83	63.90
21	23.47	21.72	113.65	0.45	71.51	39.72	348.49	50.72

说明：原文收集 95 例肝肾功能正常无发热的糖尿病住院患者，但并未给出具体的检测指标值，这里为方便介绍，根据文中提供的均数和标准差模拟产生以上 21 例数据。

主成分分析的 SAS 程序为：

```
data nkkysj14_1;
input i x1 -x8 @@ ;
cards;
1 20.18 18.77 132.34 3.23 64.92 34.89 342.53 1.19 2 23.29 24.30 138.57 2.29 51.24 47.57
367.00 59.30 3 22.44 20.18 133.99 1.45 53.45 39.96 72.27 60.60 4 22.14 20.77 126.70 1.86
72.76 37.64 283.20 61.97 5 16.81 23.95 145.45 2.77 75.84 49.97 339.38 39.65 6 17.13 20.
31 132.99 2.70 64.12 45.26 328.60 48.78 7 22.29 20.06 132.88 0.39 72.83 41.20 398.70 32.
94 8 29.24 22.41 136.95 3.19 84.82 46.24 170.90 31.15 9 22.76 18.53 112.37 1.62 70.02 50.
40 301.97 22.80 10 22.29 20.79 92.72 2.28 76.12 39.62 343.92 39.65 11 22.21 23.24 144.92
1.88 57.03 31.05 355.15 43.95 12 19.29 17.60 149.52 0.91 59.76 37.50 448.10 54.87 13 28.
50 20.15 138.62 2.58 59.80 45.92 208.91 64.65 14 20.32 19.00 116.02 1.59 67.23 38.10
369.67 69.77 15 21.83 20.07 124.53 1.28 71.71 31.61 338.32 44.80 16 17.06 18.17 170.35
0.49 57.66 38.18 320.88 33.03 17 17.99 23.45 120.88 2.50 85.35 42.40 459.46 65.16 18 27.
03 23.93 123.96 1.64 62.85 33.37 304.17 68.40 19 26.04 17.98 149.27 1.33 57.51 37.91
273.26 27.99 20 25.50 20.72 147.79 1.31 73.29 43.69 176.83 63.90 21 23.47 21.72 113.65
0.45 71.51 39.72 348.49 50.72
;
```

```
run;
ods html;
proc princomp out=a prefix=z;
    Var x1-x8; run;
proc print data=a;
Var x1-x8 z1-z8;
run;
ods html close;
```

程序说明：

通过调用 princomp 过程进行主成分分析，数据集 a 为输出数据集，其中包含了原始指标值和对应的 8 个主成分值；选项 prefix = z，表示用 z_1-z_8 命名输出的 8 个主成分变量。

主要输出结果：

(第一部分)

Correlation Matrix								
	x1	x2	x3	x4	x5	x6	x7	x8
x1	1.0000	0.1316	-.1013	0.0546	-.0105	0.0234	-.5597	0.1132
x2	0.1316	1.0000	-.1443	0.3778	0.2358	0.1678	0.0516	0.3568
x3	-.1013	-.1443	1.0000	-.1508	-.4368	-.0042	-.1565	-.0977
x4	0.0546	0.3778	-.1508	1.0000	0.2417	0.3271	-.1289	-.1451
x5	-.0105	0.2358	-.4368	0.2417	1.0000	0.2356	0.1366	-.0740
x6	0.0234	0.1678	-.0042	0.3271	0.2356	1.0000	-.1845	-.0104
x7	-.5597	0.0516	-.1565	-.1289	0.1366	-.1845	1.0000	-.0673
x8	0.1132	0.3568	-.0977	-.1451	-.0740	-.0104	-.0673	1.0000

(第二部分)

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	1.98809058	0.34141455	0.2485	0.2485
2	1.64667603	0.35914919	0.2058	0.4543
3	1.28752684	0.21602633	0.1609	0.6153
4	1.07150051	0.30271724	0.1339	0.7492
5	0.76878327	0.22714565	0.0961	0.8453
6	0.54163762	0.12060915	0.0677	0.9130
7	0.42102847	0.14627179	0.0526	0.9657
8	0.27475667		0.0343	1.0000

(第三部分)

The PRINCOMP Procedure								
Eigenvectors								
	z1	z2	z3	z4	z5	z6	z7	z8
x1	0.212100	0.574091	0.110801	-.408927	-.134885	0.225918	0.435424	0.432560
x2	0.475295	-.010173	0.343199	0.398411	-.329450	0.330385	0.255402	-.463199
x3	-.358770	0.274031	-.230530	0.572459	-.047638	0.552210	-.170721	0.281047

x4	0.477594	-.003961	-.346576	0.218916	-.507274	-.373374	-.264768	0.370152
x5	0.449453	-.329722	-.083479	-.318114	0.243124	0.572193	-.419335	0.136950
x6	0.373065	0.083849	-.380644	0.327367	0.667402	-.133078	0.371944	0.015987
x7	-.117317	-.673506	0.162844	0.137241	-.086202	0.066615	0.472282	0.502348
x8	0.131711	0.160777	0.719932	0.266859	0.318915	-.214268	-.329688	0.333129

结果及结果解释:

SAS 软件中 Princomp 过程进行主成分分析的结果主要包括一般统计描述(即原始指标的均数和标准差)、原始指标的相关系数矩阵 \mathbf{R} (见主要结果第一部分)、相关系数矩阵的特征值、贡献率及累计贡献率(见主要结果第二部分),以及特征值所对应的特征向量(见主要结果第三部分)。

本例计算结果显示 4 个大于 1 的特征根,即 $\lambda_1 = 1.988$, $\lambda_2 = 1.647$, $\lambda_3 = 1.288$, $\lambda_4 = 1.072$, 这 4 个主成分的累计贡献率为 74.92%。

第 1 主成分: $z_1 = 0.2121x_1 + 0.4753x_2 - 0.3588x_3 + 0.4776x_4 + 0.4495x_5 + 0.3731x_6 - 0.1173x_7 + 0.1317x_8$

第 2 主成分: $z_2 = 0.5741x_1 - 0.0102x_2 + 0.2740x_3 - 0.0040x_4 - 0.3297x_5 + 0.0838x_6 - 0.6735x_7 + 0.1608x_8$

第 3 主成分: $z_3 = 0.1108x_1 + 0.3432x_2 - 0.2305x_3 - 0.3466x_4 + 0.0835x_5 + 0.3806x_6 + 0.1628x_7 + 0.7199x_8$

第 4 主成分: $z_4 = -0.4089x_1 + 0.3984x_2 + 0.5725x_3 + 0.2189x_4 - 0.3181x_5 + 0.3274x_6 + 0.1372x_7 - 0.2669x_8$

接下来可以根据专业知识对这 4 个主成分进行命名。因本文数据为模拟数据,专业意义不明显,不便给出主成分的名称。

(曹文君 赵晓蒙)

参 考 文 献

[1] 葛声, 蔡东联, 唐彦, 等. 采用主成分分析法对糖尿病住院患者营养状况综合评价. 中国临床营养杂志, 2005, 13(5): 285-288.

第 15 章 探索性因子分析

多变量降维的另一典型统计分析方法是因子分析。探索性因子分析的目的也是从众多相关的指标中找出少数几个综合性指标来反映原始指标所包含的主要信息，它是用少数几个因子反映原始数据的绝大多数信息的统计分析方法。探索性因子分析的主要应用是寻找公共因子(即隐变量)和间接实现样品聚类，此法与主成分分析的最大区别是：主成分分析一步求出用原变量线性表达“主成分(即隐变量)”的表达式中的系数；而探索性因子分析通过三步才求出类似主成分分析的表达式。第一步，求出用公因子线性表达原变量的“初始因子表达式”；第二步，采取某种因子轴的旋转方法，获得旋转后的用公因子线性表达原变量的“旋转后因子表达式”；第三步，将“旋转后因子表达式”对应的系数矩阵转置，从而获得用原变量线性表达“公因子(即隐变量)”的表达式中的系数矩阵，即获得所谓的“因子得分模型”。显然，两种方法的最大区别在于是否采取上面的“第二步”！从理论上讲，经过第二步得到的结果更易于被解释、更有利于给“隐变量”命名。

本章中只介绍探索性因子分析，而不介绍验证性因子分析，以下将简称为“因子分析”。

15.1 因子分析模型

设有 p 维观测变量 $X_i (i = 1, 2, \dots, p)$ ，其均值为 $\mu_i (i = 1, 2, \dots, p)$ 。假设从 p 个变量中找到 m 个共同因子，那么因子分析可以通过下面的数学模型来表示：

$$\begin{cases} X_1 = \mu_1 + a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m + \varepsilon_1 \\ X_2 = \mu_2 + a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m + \varepsilon_2 \\ \vdots \\ X_p = \mu_p + a_{p1}f_1 + a_{p2}f_2 + \dots + a_{pm}f_m + \varepsilon_p \end{cases} \quad (15-1)$$

式(15-1)写成矩阵形式就是：

$$X = \mu + Af + \varepsilon \quad (15-2)$$

其中， f_1, f_2, \dots, f_m 为公共因子， $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ 为特殊因子，表示可以使用公共因子预测每个观测变量， $A = (a_{ij}) : p \times m$ 称为因子载荷矩阵。因子分析得到的因子载荷矩阵并不唯一，在最先得到的初始因子和原始变量的关系不清晰时，可以通过因子轴旋转得到新的潜在变量(即因子)，使新得到的因子更具代表性。常用的因子旋转方法有方差最大正交旋转和斜交旋转。其中，方差最大正交旋转的基本思想是，在保持原有公因子的正交性和公共方差总和不变的条件下，使公因子的相对载荷的方差之和最大。这样可使每个因子上的具有最大载荷的变量数最小，因此可以简化对因子的解释。斜交旋转适用于大数据集的因子分析，经斜交旋转后，各因子载荷发生较大变化，各因子间不再相互独立，各因子对各变量的贡献的总和也发生了改变。

15.2 因子分析的步骤

1. 判别变量是否适合做因子分析

常用的检验方法包括 Bartlett 球形检验、KMO(Kaiser-Meyer-Olkin)检验。

Bartlett 球形检验是通过相关矩阵检验原始变量间的相关性,它假定变量相关矩阵为单位矩阵,进而检验实际相关矩阵与单位矩阵的统计差异,如果差异有统计学意义,则拒绝相关矩阵为单位阵的假设,即认为原变量间的相关性有统计学意义,适合做因子分析,否则不适合做因子分析。

KMO (Kaiser-Meyer-Olkin) 统计量用于检验变量间的偏相关性是否足够小,是变量相关系数与偏相关系数的一个相对指数,计算公式为:

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}^2}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} p_{ij}^2} \quad (15-3)$$

式中, r_{ij} 表示两变量间的相关系数, p_{ij} 为两变量间的偏相关系数。偏相关系数为将其他变量控制在特定水平条件下,所计算的某两个变量之间的相关系数。KMO 统计量取值范围在 0 和 1 之间。与 Bartlett 球形检验的本质一样,如原始变量间相互作用较大,简单相关系数相对较大,则变量间的偏相关系数就会相对较小,由式(15-3)可知, KMO 值就越大,适合做因子分析,反之不适合做因子分析。Kaiser 提供的判别标准是:当 $KMO > 0.9$ 时,非常适合;当 $0.8 < KMO < 0.9$ 时,适合;当 $0.7 < KMO < 0.8$ 时,一般;当 $0.6 < KMO < 0.7$ 时,不太适合;当 $KMO < 0.5$ 时,不适合。

2. 提取因子

因子分析的一个关键环节是确定提取的公因子及其个数。常用的方法有“主成分分析法”、“极大似然估计法”、“主轴因子法”等,其中以“主成分分析法”的使用最为普遍。应保留的因子个数要视具体情况而定,一般以因子累计贡献率达 75% 以上为好,但这个标准并不统一,有学者认为大于 80% 或 85% 为好。因子贡献率反映每个公共因子解释原始数据的能力,它等于与该因子有关的因子载荷的平方和,表示每个公共因子所解释的方差占有变量总方差的比例。此外,还可以通过制定保留的最大因子数目或制定保留的最小特征根,同主成分分析一样,一般取特征根大于 1。也可以采用碎石图来直观判别需要提取的因子数,一般选择拐点之前的特征根数目为提取公因子的数目。

3. 因子的命名和解释

因子的命名和解释通常需要具备一定的专业基础知识,要对每个因子赋予一个具有实际意义的名称。如果实际操作中公共因子不易解释,则可通过旋转公共因子,并对其进行命名。公共因子的命名要根据因子变量与原始变量之间的关系,也就是要观察旋转后的因子载荷矩阵中需要命名的公共因子能够解释的变量,以确定公共因子的特征,最后得到符合实际的名称。

4. 计算因子得分

计算因子得分就是首先根据因子模型建立每个因子的回归方程,然后将原变量得分代入计算该变量的因子得分。常用的方法包括加权最小二乘法(也称 Bartlett 因子得分)和回归法等。

15.3 实例分析

【例 15-1】 实测得到性别(1 - 男; 2 - 女) X_1 、月龄 X_2 、身长 X_3 、体重 X_4 、胸围 X_5 及心像面积 X_6 这 6 项指标,为了从这 6 项指标中找出潜在的共同变量,测定 13 名儿童的 6 项指标值(见表 15-1)。对表 15-1 的数据采用因子分析。

表 15-1 6 项指标测定值

<i>i</i>	<i>X</i> ₁	<i>X</i> ₂	<i>X</i> ₃	<i>X</i> ₄	<i>X</i> ₅	<i>X</i> ₆
1	1	32	95.5	14.0	53.5	49.64
2	1	35	92.0	13.0	52.0	41.61
3	1	33	89.0	12.5	53.5	35.81
4	1	176	168.0	53.5	82.0	100.14
5	1	96	117.0	19.7	56.0	67.20
6	1	96	113.0	18.1	55.0	66.00
7	1	96	122.0	21.6	57.3	58.00
8	2	30	91.0	11.0	48.0	35.39
9	2	33	91.0	11.5	47.0	44.98
10	2	33	91.0	12.5	50.0	29.51
11	2	176	156.0	55.0	83.0	94.66
12	2	178	163.0	54.0	79.0	87.42
13	2	84	130.0	25.0	58.0	62.00

因子分析的 SAS 程序为：

```
data nkkysj15_1;
input i X1 - X6 @@ ;
cards;
1 1 32 95.5 14.0 53.5 49.64 2 1 35 92.0 13.0 52.0 41.61 3 1 33 89.0 12.5 53.5 35.81
4 1 176 168.0 53.5 82.0 100.14 5 1 96 117.0 19.7 56.0 67.20 6 1 96 113.0 18.1 55.0
66.00 7 1 96 122.0 21.6 57.3 58.00 8 2 30 91.0 11.0 48.0 35.39 9 2 33 91.0 11.5 47.
0 44.98 10 2 33 91.0 12.5 50.0 29.51 11 2 176 156.0 55.0 83.0 94.66 12 2 178 163.0
54.0 79.0 87.42 13 2 84 130.0 25.0 58.0 62.00
;
run;
proc factor method=prin rotate=varimax score outstat=a;
var X1 - X6;
run;
proc score data=nkkysj15_1 out=aa;
proc print data=aa;
var factor1 - factor2;
run;
```

程序说明：

调用 factor 过程进行因子分析。选项 method = prin 表示采用主成分法提取因子；选项 rotate = varimax 表示做最大方差旋转；选项 score 表示计算因子得分系数，并将结果保存在数据集 a 中。score 过程用于计算每个研究对象的因子得分，并将输出结果保存在数据集 aa 中。

主要输出结果：

（第一部分）

The FACTOR Procedure	
Input Data Type	Raw Data
Number of Records Read	13
Number of Records Used	13
N for Significance Tests	13

The FACTOR Procedure
Initial Factor Method: Principal Components
Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 6 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	4.84195107	3.82901145	0.8070	0.8070
2	1.01293962	0.91913569	0.1688	0.9758
3	0.09380394	0.06231117	0.0156	0.9914
4	0.03149277	0.01310174	0.0052	0.9967
5	0.01839103	0.01696945	0.0031	0.9998
6	0.00142158		0.0002	1.0000

2 factors will be retained by the MINEIGEN criterion.

(第二部分)

Factor Pattern		
	Factor1	Factor2
X1	0.11797	0.99234
X2	0.98761	-0.04521
X3	0.98831	-0.00183
X4	0.98830	0.07992
X5	0.97955	-0.01431
X6	0.96936	-0.13986

Variance Explained by Each Factor

Factor1	Factor2
4.8419511	1.0129396

Final Communality Estimates: Total = 5.854891

X1	X2	X3	X4	X5	X6
0.99865728		0.97741643	0.97675960	0.98311715	0.95971400
					0.95922622

(第三部分)

Rotation Method: Varimax
Orthogonal Transformation Matrix

	1	2
1	0.99733	0.07302
2	-0.07302	0.99733

Rotated Factor Pattern

	Factor1	Factor2
X1	0.04520	0.99831
X2	0.98827	0.02702
X3	0.98581	0.07034

	X4	0.97982	0.15187		
	X5	0.97798	0.05725		
	X6	0.97699	-0.06871		
Variance Explained by Each Factor					
	Factor1	Factor2			
	4.8215375	1.0333532			
Final Communality Estimates: Total = 5.854891					
X1	X2	X3	X4	X5	X6
0.99865728	0.97741643	0.97675960	0.98311715	0.95971400	0.95922622

(第四部分)

The FACTOR Procedure					
Rotation Method: Varimax					
Scoring Coefficients Estimated by Regression					
Squared Multiple Correlations of the Variables with Each Factor					
	Factor1	Factor2			
	1.0000000	1.0000000			
Standardized Scoring Coefficients					
	Factor1	Factor2			
X1	-0.04723	0.97883			
X2	0.20668	-0.02962			
X3	0.20370	0.01310			
X4	0.19781	0.09359			
X5	0.20280	0.00068			
X6	0.20975	-0.12309			

(第五部分)

Obs	Factor1	Factor2
1	-0.59559	-0.86032
2	-0.71713	-0.82623
3	-0.77947	-0.79839
4	1.76805	-0.95314
5	0.04056	-0.94506
6	-0.03210	-0.94918
7	0.03421	-0.88363
8	-0.97476	1.08428
9	-0.88756	1.03450
10	-0.96882	1.12222
11	1.57710	0.96521
12	1.49264	1.00016
13	0.04285	1.00957

结果解释：

第一部分为 factor 过程用主成分分析法提取初始公因子情况。具体地说，输出的内容是 6 个变量对应的相关矩阵的特征值、相邻两个特征值的差值、每个特征值的贡献率和全部特征值的贡献率由大到小排序后计算出来的累计贡献率。在本例中，前 2 个特征值大于 1，累积方差贡献率为 97.58%，解释原变量的总信息量很大，因此，保留前 2 个公共因子。

第二部分列出未经旋转的所有变量在前 2 个公因子上的载荷矩阵，前 2 个公因子能解释的方差，变量 X1 - X6 对最终共性方差的估计值。

第三部分给出最大方差旋转的正交矩阵，因子旋转后 6 个变量在前 2 个公因子上的载荷矩阵，前 2 个公因子能解释的方差，变量 X1 - X6 对最终共性方差的估计值；因子 Factor1 在 X_2 (月龄)、 X_3 (身高)、 X_4 (体重)、 X_5 (胸围)、 X_6 (心像面积) 上载荷较大，分别为 0.98827、0.98581、0.97982、0.97798、0.97699。因子 Factor2 在 X_1 (性别) 上载荷较大，为 0.99831。结合专业知识，根据因子载荷大小将 Factor1 命名为“发育因子”，Factor2 命名为“性别因子”。

第四部分为标准化的因子得分系数矩阵，由因子得分系数矩阵，可以将公因子表示为各变量的线性组合，得到的因子得分函数为：

$$\begin{cases} f_1 = -0.04723x_1 + 0.20668x_2 + 0.20370x_3 + 0.19781x_4 + 0.20280x_5 + 0.20975x_6 \\ f_2 = 0.97883x_1 - 0.02962x_2 + 0.01310x_3 + 0.09359x_4 + 0.00068x_5 - 0.12309x_6 \end{cases}$$

第五部分为样本在前 2 个公因子上的得分。

(曹文君 石福艳)

第 16 章 遗传数据统计分析

如何科学、合理地分析遗传资料？显然，常规的统计分析方法不能完全胜任。因为这是一个特殊的医学研究分支，有很多专门的概念和计算方法。SAS/Genetics 模块中的 allele、haplotype、htsnp、casecontrol、family 等过程可以直接用来解决一些常见的遗传资料的分析问题，但有些较复杂的遗传资料需借助 SAS 软件中其他模块提供的计算过程方可得到圆满解决。本章将介绍与遗传统计分析有关的概念、计算原理、用 SAS 实现统计计算的具体方法及结果解释。

16.1 基因、基因型频率测定与 Hardy-Weinberg 平衡定律的验证

哈代 - 温伯格 (Hardy-Weinberg) 平衡定律是在一种非常理想化的状态，有很多假设条件作为前提，如随机婚配，所研究的生物是二倍体，繁殖方法为两性繁殖，代与代之间没有重叠，男性和女性等位基因概率相同，群体的个数是无数的，突变可以被忽略，等等。在实际应用中，多数情况下上述条件都满足或者近似满足，人们也经常假定哈代 - 温伯格平衡定律成立，但仍有必要对这一假定进行统计学检验。

哈代 - 温伯格平衡定律：群体(继代)随机交配，在不产生选择、突变和迁移的情况下，基因频率与基因型频率每代保持不变，合子系列频率等于配子系列频率的二项式平方。写成：

$$(Ap_A + ap_a)^2 = AA p_A^2 + Aa 2p_A p_a + aa p_a^2 \quad (16-1)$$

对于一个给定的遗传资料，为检验其是否符合哈代 - 温伯格平衡定律，可以使用拟合优度检验来回答，其相应的 Pearson χ^2 统计量为：

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad (16-2)$$

上式是对某一给定基因座位的所有基因型求和， O (observed) 代表每一基因型的观察数目， E (expected) 代表每一基因型在哈代 - 温伯格平衡定律成立假定下期望数目。考虑一个有两个等位基因的基因座位，令 n_{AA} 、 n_{Aa} 和 n_{aa} 分别表示具有基因型 AA 、 Aa 和 aa 的个体的数目，记个体总数 $n = n_{AA} + n_{Aa} + n_{aa}$ 。在零假设 H_0 (哈代 - 温伯格平衡定律) 下， (n_{AA}, n_{Aa}, n_{aa}) 服从三项分布，分布参数为 $(n, p^2, 2pq, q^2)$ ，因此三种基因型的期望数目分别为 np^2 、 $2npq$ 和 nq^2 。等位基因 A 出现的概率 p 和等位基因 a 出现的概率 q 可以分别估计为：

$$\hat{p} = \frac{2n_{AA} + n_{Aa}}{2n}, \hat{q} = \frac{2n_{aa} + n_{Aa}}{2n} \quad (16-3)$$

从而得到三种基因型出现的期望数目的估计为：

$$\hat{E}_{AA} = n\hat{p}^2, \hat{E}_{Aa} = 2n\hat{p}\hat{q}, \hat{E}_{aa} = n\hat{q}^2 \quad (16-4)$$

由此可以得到 Pearson χ^2 统计量为：

$$\chi^2 = \frac{(n_{AA} - \hat{E}_{AA})^2}{\hat{E}_{AA}} + \frac{(n_{Aa} - \hat{E}_{Aa})^2}{\hat{E}_{Aa}} + \frac{(n_{aa} - \hat{E}_{aa})^2}{\hat{E}_{aa}} \quad (16-5)$$

此检验统计量的自由度为 1。如果 χ^2 的样本观测值大于 $\chi^2_{\alpha}(1)$ ，则可拒绝零假设 H_0 (即该资料符合哈代 - 温伯格平衡定律)，从而可以认为基因型不具有哈代 - 温伯格比例，这里的 α 是检验的显著

性水平，而 $\chi^2_{\alpha}(1)$ 是自由度为 1 且显著性水平为 α 时的 χ^2 值。

【例 16-1】 脂联素是脂肪组织分泌的一种细胞因子，研究显示糖尿病、胰岛素抵抗、肥胖患者有较低的脂联素水平，某研究机构筛查了 65 例正常对照和 78 例代谢综合征患者的脂联素外显子 2Gly15Gly 基因型，其中正常对照组基因型为 TT, TG, GG 的人数分别为 34, 26, 5；代谢综合征患者组基因型为 TT, TG, GG 的人数分别为 36, 33, 9。试计算两组人群脂联素外显子 2Gly15Gly 基因型频率，并验证两组人群基因型频率是否符合哈代 – 温伯格平衡定律。

例 16-1 资料的程序名为 SAS16_1.SAS。

程 序	说 明
<pre>data nkkysj16_1; input group \$ (a1 - a2) (\$) f@@; cards; U T T 34 U T G 26 U G G 5 A T T 36 A T G 33 A G G 9 ; run; data two; set nkkysj16_1; do i = 1 to f; output two; end;run; Ods html; proc sort data = two; by group; proc allele data = two prefix = marker perms = 10000 boot = 1000 seed = 123; var a1 a2; by group; run; Ods html close;</pre>	<p>建立数据集</p> <p>输入变量 a1 a2 为 2Gly15Gly 基因型。U 代表正常对照组，T 代表代谢综合征组。</p> <p>调用 allele 过程进行计算，并使用 bootstrap (自助法)的方法估计置信区间，seed 为随机种子数。</p>

主要分析结果及解释如下。

The ALLELE Procedure										
group = A										
Marker Summary										
Locus	Number of Indiv	Number of Alleles	Polymorph Info Content	Heterozygosity	Allelic Diversity	Test for HWE				
						Chi-Square	DF	Pr >	ChiSq	Prob Exact
marker1	78	2	0.3432	0.4231	0.4401	0.1166	1	0.7328		0.7985

这部分结果给出了病例组人群哈代 – 温伯格平衡定律的验证结果，“Test for HWE”中计算结果显示卡方值为 0.1166, $P=0.7328>0.05$ 。说明病例组人群脂联素外显子 2Gly15Gly 基因型频率均符合哈代 – 温伯格平衡定律。

Allele Frequencies							
Locus	Allele	Count	Frequency	Standard Error	95% Confidence Limits		
marker1	G	51	0.3269	0.0383	0.2564	0.4038	
	T	105	0.6731	0.0383	0.5962	0.7436	

这部分输出结果给出了病例组人群脂联素外显子 2Gly15Gly 的基因频率及标准误、置信区间。

Genotype Frequencies							
Locus	Genotype	Count	Frequency	HWD Coeff	Standard Error	95% Confidence Limits	
marker1	G/G	9	0.1154	0.0085	0.0252	-0.0405	0.0552
	G/T	33	0.4231	0.0085	0.0252	-0.0405	0.0552
	T/T	36	0.4615	0.0085	0.0252	-0.0405	0.0552

这部分输出结果给出了病例组人群脂联素外显子 2Gly15Gly 的基因型频率及标准误、置信区间。

The ALLELE Procedure										
group = U										
Marker Summary										
Locus	Number of Indiv	Number of Alleles	Polymorph Info Content	Heterozygosity	Allelic Diversity	Chi-Square	DF	Pr > ChiSq	Prob Exact	
marker1	65	2	0.3203	0.4000	0.4005	0.0001	1	0.9924	1.0000	

这部分结果给出了正常对照组人群哈代 – 温伯格定律的验证结果，“Test for HWE”中计算结果显示卡方值为 0.0001， $P=0.9924>0.05$ 。说明对照组人群脂联素外显子 2Gly15Gly 基因型频率均符合哈代 – 温伯格定律。

Allele Frequencies							
Locus	Allele	Count	Frequency	Standard Error	95% Confidence Limits		
marker1	G	36	0.2769	0.0393	0.2077	0.3538	
	T	94	0.7231	0.0393	0.6462	0.7923	

这部分输出结果给出了正常对照组人群脂联素外显子 2Gly15Gly 的基因频率及标准误、置信区间。

Genotype Frequencies							
Locus	Genotype	Count	Frequency	HWD Coeff	Standard Error	95% Confidence Limits	
marker1	G/G	5	0.0769	0.0002	0.0249	-0.0502	0.0530
	G/T	26	0.4000	0.0002	0.0249	-0.0502	0.0530
	T/T	34	0.5231	0.0002	0.0249	-0.0502	0.0530

这部分输出结果给出了正常对照组人群脂联素外显子 2Gly15Gly 的基因型频率及标准误、置信区间。

专业结论：依据哈代 – 温伯格定律进行计算。可得到两组人群脂联素外显子 2Gly15Gly 基因型频率符合哈代 – 温伯格定律。

16.2 连锁不平衡与单体型分析

16.2.1 概述

连锁不平衡(linkage disequilibrium, LD)是指相邻基因座位上等位基因的非随机性相关。当位于某一基因座位上的等位基因与同一条染色体另一基因座位上的某等位基因同时出现的概率高于或低于人群中因随机分布而使两位点同时出现的概率时,就称这两个位点处于连锁不平衡状态。

单体型是指一条染色体上紧密连锁的多个等位基因的线性排列。SNP 单体型就是不同 SNPs 位点上核苷酸碱基的线性排列,每一种线性排列称为一种 SNP 单体型。如果在某一段 DNA 片段上发现了 10 个 SNPs,理论上可能存在 $1024(2^{10})$ 种单体型,但由于连锁不平衡的存在,实际发现的单体型数目往往远远少于理论上的值。在对人类遗传疾病的研究中,科学家们发现许多人类基因疾病往往不是由单个 SNP 位点引起的,而是由若干个位点上的 SNPs 联合作用导致的,即 SNP 单体型控制疾病的发生,这就是研究单体型意义所在。在对 SNP 单体型数据的处理中,我们需要估计单体型概率。

16.2.2 最大似然法估计单体型概率

如果有 L 个 SNPs, 则共有 $K = 2^L$ 种可能的单体型。记所有可能的单体型为 h_1, \dots, h_K , 其在人群中相应的频率为 $P = (p_1, \dots, p_K)'$, 我们的目的是要估计 P 。假定通过观测得到 n 个个体的基因型 $G_i, i = 1, \dots, n$, 而这些个体的单体型对为 $H_i = (H_i^{(1)}, H_i^{(2)})$ (本节提及的单体型对都是计较顺序的, 比如, $H_i^{(1)}$ 来自父亲而 $H_i^{(2)}$ 来自母亲), 则完全数据 $H_i, i = 1, \dots, n$ 的似然函数为:

$$L_C(P) = \prod_{i=1}^n P(H_i) = \prod_{i=1}^n \prod_{K=1}^K \prod_{L=1}^K (p_K p_L)^{I(H_i = (h_K, h_L))} \quad (16-6)$$

这里用到了哈代 – 温伯格平衡定律成立的假定, 所以计序单体型对 (ordered haplotype pair) (h_K, h_L) 的概率为 $p_K p_L$ 。观测数据的似然函数为:

$$L_O(P) = \prod_{i=1}^n P(G_i) = \prod_{i=1}^n \sum_{H \in S(G_i)} P(H) = \prod_{i=1}^n \sum_{H \in S(G_i)} \prod_{K=1}^K \prod_{L=1}^K (p_K p_L)^{I(H = (h_K, h_L))} \quad (16-7)$$

其中, $S(g)$ 表示所有与 g 相容的单体型对组成的集合。例如, 当 SNPs 的个数为 $L = 2$ 时, 基因型 $(1, 1)$ 对应于 4 种单体型对: 01/10, 00/11, 10/01, 11/00。这里 0 表示特定位置的一个碱基, 而 1 表示碱基对中的另外一个碱基。注意到某位点的基因型正好可以用两个等位基因代码 (0 或 1) 的和来表示, 即 0 表示基因型 0/0, 1 表示基因型 0/1, 2 表示基因型 1/1。

从理论上讲, 我们可以基于上述观测似然函数求解最大似然估计 (maximum likelihood estimate, MLE), 但是当 SNPs 的个数 L 稍微增大一点时, 求解 MLE 所需的计算量非常大, 下面给出 E-M 算法求解。

16.2.3 E-M 算法估计单体型概率及对连锁不平衡进行假设检验

对于 2 个位点的单体型, 可以使用最大似然估计的方法求得单体型概率, 但对于多于 2 个位点的单体型而言, 一般无法得到单体型概率最大似然估计的解析解。在假定哈代 – 温伯格平衡的前提下, 可以用最大数学期望值算法 (expectation-maximization, E-M) 来得到单体型频率的最大似然估计。E-M 算法是一种迭代算法, 每一次迭代由两步组成, 即 E 步 (求期望) 和 M 步 (极大化)。对于

一个群体有 n 例个体，假定第 i 例个体的基因型为 G_i ， G_i 在群体中的频率为 P_i ，E 步求得对数似然函数为：

$$\log L = \sum_{i=1}^n \log P_i \tag{16-8}$$

其中，

$$P_i = \sum_{j \in H_i} f_j^{c_i} \tag{16-9}$$

h_j 表示第 j 个可能的单体型， H_i 表示这些个体 ($i=1, 2, 3, \dots, n$) 存在的单体型对，即 $(h_j, h_j^{c_i})$ ， f_j 表示该单体型 h_j 频率， $f_j^{c_i}$ 为构成基因型 G_i 的单体型 $h_j^{c_i}$ 的频率。在第一次运算中，单体型频率 f_j 是由初始值给定的，初始值是随机的，也可以用其他方法计算得出。然后，我们根据两次迭代之间的收敛性来确定是否继续迭代，如果程序继续，则计算出 M 步(极大化)中新的 f_j 值：

$$f_j = \frac{1}{2n} \sum_{i=1}^n \frac{m_{ij} f_j^{c_i}}{P_i} \tag{16-10}$$

其中， m_{ij} 是基因型 G_i 中单体型 h_j 出现的次数，如果是完全吻合，则该个体有 2 种单体型，即 m_{ij} 取 2；如果为杂合型， m_{ij} 取 1 或 0。此时求得新的 f_j 值，一次迭代计算结束，将 f_j 值代入 E 步，继续计算，每进行一次迭代，对数似然函数就会增加，直到达到最大值迭代运算结束。

对于多于 2 个位点的单体型概率估计而言，原则上可以运用 E-M 算法计算任意多个位点的单体型概率。但是随着生物检测技术的增强，研究位点不断增多，计算时间急剧增加以至于使用 E-M 算法不现实，有的学者提出了一些改进算法，例如分割－捆绑算法(partition-ligation algorithm)，将大量的 SNP 位点分割成“单位块”，再利用 E-M 算法估计每个“单位块”的单体型概率，每个“单位块”都会得出若干个非零单体型，再将相邻 2 个“单位块”进行合并，即将 2 个“单位块”中非零单体型数目相乘积从而得到新的单体型，如此计算大大减少了计算量。对于多于 2 个位点而 SNP 位点个数又不是特别多的单体型概率估计运用 EM 算法可以得到较为理想的结果，显示了该算法在 SNP 位点单体型概率估计中的实用性。

对于连锁不平衡的检验，基于如下运算思想。

原假设 H_0 ：各个位点之间不存在连锁不平衡，此时单体型频率即等位基因频率运算直接得到，求出对数似然函数，记为 $\log L_0$ ，自由度为 $df_0 = \sum_{i=1}^m (k_i - 1)$ 。

备择假设 H_1 ：各个位点之间存在连锁不平衡，利用 E-M 算法求出对数似然函数，记为 $\log L_1$ ，自由度为 $df_1 = \text{单体型数目} - 1$ 。

对其进行假设检验，有：

$$2(\log L_1 - \log L_0) \sim \chi^2 \tag{16-11}$$

自由度： $df = df_1 - df_0$

【例 16-2】 某研究选取血管紧张素原(angiotensinogen, AGT) 基因启动子区 -217、-152、-20、-6、内含子 1 的 +31，共 5 个位点，对 100 例的正常人进行了 SNP 位点基因型检测，数据如表 16-1 所示，计算两两位点间的连锁不平衡参数，并验证 5 个位点是否存在连锁不平衡关系，求得可能存在的单体型，并计算单体型概率。

表 16-1 5 个 SNP 位点的基因型数据

编号	G217-A	G152-A	A-20C	A-6G	C+31T
1	A A	A A	A A	A A	C C
2	A A	A A	A A	A A	C C
⋮	⋮	⋮	⋮	⋮	⋮
100	G G	A A	C C	A G	T T

例 16-2 资料的程序名为 SAS16_2. SAS。

程 序	说 明
<pre>data nkkysj16_2; input a1 - a10 @@ ; cards; 1 2 2 1 1 2 2 1 2 2 2 ; run; Ods html; proc allele data = nkkysj16_2 prefix = SNP haplo = est corrcoeff dprime; var a1 - a10; run; proc haplotype data = nkkysj16_2 ld; var a1 - a10; run; Ods html close;</pre>	<p>建立数据集</p> <p>以下是输入数据(原始数据较多,为节省篇幅,此处省略部分数据)。</p> <p>调用 allele 过程进行计算,对 SNP 位点进行两位点间连锁不平衡参数计算,显示相关系数。</p> <p>调用 HAPLOTYPE 过程,利用 EM 算法求得各个单体型频率,并验证各个位点间是否存在连锁不平衡。“LD”选项即进行连锁不平衡检验,如果只求得单体型频率,可更换为“EST = EM”。</p>

主要分析结果及解释如下。

Linkage Disequilibrium Measures						
Locus1	Locus2	Haplotype	Frequency	LD Coeff	Corr Coeff	Lewontin's D'
SNP1	SNP2	1 - 1	0.8200	-0.0065	-0.0887	-1.0000
SNP1	SNP2	1 - 2	0.0500	0.0065	0.0887	1.0000
SNP4	SNP5	1 - 2	0.1585	-0.0159	-0.0911	-0.0911
⋮	⋮	⋮	⋮	⋮	⋮	⋮
SNP4	SNP5	2 - 2	0.0665	0.0159	0.0911	0.0911

这部分给出了 5 个 SNP 位点两两连锁不平衡参数(LD COEFF)的计算结果,以及相关系数 r 值、测度参数 D'值。

Haplotype Frequencies						
Number	Haplotype	H0 Freq	H1 Freq	Standard Error	95% Confidence Limits	
1	1 - 1 - 1 - 1 - 1	0.42692	0.54720	0.03529	0.47804	0.61636
2	1 - 1 - 1 - 1 - 2	0.12394	0.09570	0.02085	0.05483	0.13657
3	1 - 1 - 1 - 2 - 1	0.12394	0.12392	0.02336	0.07814	0.16970
4	1 - 1 - 1 - 2 - 2	0.03598	0.01180	0.00765	0.00000	0.02680
5	1 - 1 - 2 - 1 - 1	0.06950	0.02734	0.01156	0.00468	0.05000
6	1 - 1 - 2 - 1 - 2	0.02018	0.00516	0.00508	0.00000	0.01511

7	1 - 1 - 2 - 2 - 1	0.02018	0.00888	0.00665	0.00000	0.02191
8	1 - 1 - 2 - 2 - 2	0.00586	0.00000	0.00000	0.00000	0.00000
9	1 - 2 - 1 - 1 - 1	0.02247	0.00567	0.00532	0.00000	0.01611
10	1 - 2 - 1 - 1 - 2	0.00652	0.04341	0.01445	0.01510	0.07172
11	1 - 2 - 1 - 2 - 1	0.00652	0.00000	0.00000	0.00000	0.00000
12	1 - 2 - 1 - 2 - 2	0.00189	0.00092	0.00215	0.00000	0.00513
13	2 - 1 - 1 - 1 - 1	0.06379	0.01625	0.00896	0.00000	0.03381
14	2 - 1 - 1 - 1 - 2	0.01852	0.00438	0.00468	0.00000	0.01355
15	2 - 1 - 1 - 2 - 1	0.01852	0.00000	0.00001	0.00000	0.00001
16	2 - 1 - 1 - 2 - 2	0.00538	0.01075	0.00731	0.00000	0.02508
17	2 - 1 - 2 - 1 - 1	0.01038	0.01435	0.00843	0.00000	0.03087
18	2 - 1 - 2 - 1 - 2	0.00301	0.01555	0.00877	0.00000	0.03273
19	2 - 1 - 2 - 2 - 1	0.00301	0.03139	0.01236	0.00716	0.05561
20	2 - 1 - 2 - 2 - 2	0.00088	0.03735	0.01344	0.01100	0.06369

这部分输出结果中“H1 Freq”一列对应的结果为各种可能的单体型频率，1 - 1 - 1 - 1 - 1 和 1 - 1 - 1 - 2 - 1 是最有可能存在的两种单体型。

Test for Allelic Associations				
Hypothesis	DF	LogLike	Chi-Square	Pr > ChiSq
H0: No Association	5	-329.44726		
H1: Allelic Associations	19	-282.76008	93.3744	<.0001

这部分输出结果是对 5 个位点是否存在连锁不平衡关系进行假设检验。结果为 $\chi^2 = 93.3744$ ， $P < 0.0001$ ，说明了 5 个位点之间存在连锁不平衡关系。

专业结论：5 个位点存在较强的连锁不平衡关系，1 - 1 - 1 - 1 - 1 和 1 - 1 - 1 - 2 - 1 是最有可能存在的两种单体型。

16.3 多位点基因型与疾病关联分析

对于疾病性状与多个位点之间的关联分析，SAS 运用了 χ^2 检验对其进行计算，见表 16-2。

“单体型 1”是指全部样本个体经过 EM 算法运算得到最有可能(单体型概率值最高)的一种单体型，“单体型 2”指其他单体型。 $T = 2n = t_1 + t_2 = h_1 + h_2$ ， n 为个体数。 c_{ij} 表示每个格子相应的单体型例数，但是很多个体的单体型是无法直接观察到的，所以 c_{ij} 可以使用单体型频率直接估计，所得结果可能不是整数。 h_1 、 h_2 不是通常我们认定的每一列直接求和，而是 $h_1 = T \times f_j$ ， $h_2 = T \times (1 - f_j)$ ， f_j 表示全部个体计算得到的“单体型 1”的频率。该数据服从自由度为 1 的 χ^2 分布。

表 16-2 疾病性状与多位点单体型关联表

疾病性状	单体型数量		合计
	单体型：	单体型 1 单体型 2	
正常		c_{11} c_{12}	t_1
患病		c_{21} c_{22}	t_2
合计		h_1 h_2	T

【例 16-3】 某研究探讨 β_2 肾上腺素能受体基因(β_2 AR)的多肽性与中国人哮喘易感性的连锁关系及对疾病调节作用的重要性。用 PCR/RFLP 法检测 β_2 AR 基因编码区第 16、27、164 位多肽性

位点及核苷酸 523(C – A)位点(4 个 SNP 位点), 对正常对照组 30 例、哮喘病患儿 30 例血液样品测定结果, 分别见表 16-3 和表 16-4, 试对此资料进行分析。

表 16-3 正常对照组 30 例 4 位点基因型

受试者编号	性别 (M/F)	年龄	A16G	Gln27Glu	T164I	A523C
1	M	5	A A	G G	T T	C C
2	M	6	A A	G G	T T	C C
⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	M	44	A G	G G	T T	A A

注: 实际计算中应增大样本含量。

表 16-4 30 例患儿 4 位点基因型

受试者编号	性别 (M/F)	年龄	A16G	Gln27Glu	T164I	A523C
1	M	8	A A	G G	T T	C C
2	M	5.5	A A	G G	T T	A C
⋮	⋮	⋮	⋮	⋮	⋮	⋮
30	M	11	A A	G G	T T	C C

注: 实际计算中应增大样本含量。

例 16-3 资料的程序名为 SAS16_3.SAS。

程 序	说 明
<pre>DATA nkkysj16_3; INPUT disease \$(A1 – A8)\$; CARDS; 0 A A G G T T C C 0 A A G G T T C C 0 A A G G T T C C 1 A G G G T T A C 1 A A G G T T C C ; run; ods html; proc haplotype data = nkkysj16_3 noprint; var a1 – a8; trait disease; run; ods html close;</pre>	<p>建立数据集</p> <p>以下是输入变量, disease 变量输入 0(无病)、1(有病), 基因型为 8 个变量, 每对等位基因为两个变量。</p> <p>调用 HAPLOTYPE 过程, 求得 4 个位点与哮喘病之间是否存在关联。</p>

主要分析结果及解释如下。

The HAPLOTYPE Procedure

Test for Marker – Trait Association

Trait Number	Trait Value	Num Obs	DF	LogLike	Chi – Square	Pr > ChiSq
1	0	30	6	–54.71294		

2	1	30	13	-73.41111		
	Combined	60	13	-134.99582	13.7435	0.0326

由输出结果可知, $\chi^2 = 13.7435$, $P = 0.0326 < 0.05$, 说明 4 个位点基因型的不同与疾病的发生之间是存在一定关系的。

专业结论: 4 个位点基因型的不同与疾病的发生之间存在关联, 不同人群中 4 个位点基因型频率的改变可能会致病。

16.4 标签 SNP 的确认与 SAS 程序

如果单体型被确认, 可以检查那些在单体型中特异的不同于普通单体型 SNPs, 这些特异性的 SNPs 被称为“单体型标签 SNPs (haplotype tag SNPs, htSNPs)”, htSNPs 的意义——只用少量的 htSNPs 就能代表单体型区块内绝大多数的常见单体型种类。所以 htSNPs 成为寻找致病基因的一条捷径。如果某个单体型区块在特定疾病人群中更为普遍, 那么在此单体型区块内可找到疾病相关的基因。SAS/Genetics 中提供了 HTSNP 过程对 htSNPs 的选择进行计算。

挑选标签 SNPs 有一些可行的运算方法, SAS 采用 Stram(2003) 给出的计算方法对其实现计算。设某个体的基因型为 G , 单体型对为 H , $H = (H_1, H_2)$, 对于给定的单体型 h , 用 $\delta_h(H)$ 表示 H_1 和 H_2 中等于 h 的个数。由于 H 通常观测不到而 G 能被观测到, 因此可以用 $E[\delta_h(H) | G]$ 来作为 $\delta_h(H)$ 的预测, 预测的精度可以用两者方差之比来度量, 该度量可以理解为线性回归中的复相关系数 (squared correlation), 记为 R_h^2 , 我们称其为 h 的确定度。

假定随机婚配 (位点基因型符合哈代 - 温伯格平衡定律), 则 $\delta_h(H)$ 的方差为:

$$\text{Var}(\delta_h(H)) = \text{Var}(I(H_1 = h)) + \text{Var}(I(H_2 = h)) = 2p_h(1 - p_h)$$

计算得到:

$$E[\delta_h(H) | G] = \frac{\sum_{H \in S(G)} \delta_h(H) P(H)}{\sum_{H \in S(G)} P(H)} \tag{16-12}$$

并有:

$$R_h^2 = \frac{\text{Var}[E\{\delta_h(H) | G\}]}{2p_h(1 - p_h)} \tag{16-13}$$

由定义, 对于给定的单体型分布, 我们可以算得 R_h^2 值。确定度高的一组 SNPs 说明代表性比较好, 即为我们想求得的标签 SNPs。

【例 16-4】 某研究探讨中国人白细胞介素 (IL) - 13 基因单核苷酸多态性与慢性阻塞性肺疾病之间的关系。后续表述为该实验的一部分内容, 调查了 34 个健康吸烟人 (IL) - 13 基因三个 SNP 位点, 分别为: 1103C/T、4257G/A、4738G/A, 数据已在 SAS 程序数据步中列出, 试计算得到 SNP 单体型及概率, 并求得 htSNPs。

例 16-4 资料的程序名为 SAS16_4. SAS。

程 序	说 明
DATA nkksj16_4; INPUT ID(M1 - M3)(\$) @@ ; DATALINES;	建立数据集 以下是输入变量, M1 - M3 代表输入三个位点基因型, 每个位点的两个等位基因之间用斜线连接。

续表

程 序								说 明	
1	C/C	G/G	G/G	18	C/T	G/G	G/G		
2	C/C	G/G	G/A	19	C/T	G/G	G/G		
3	C/C	G/G	G/G	20	C/T	G/G	G/G		
4	C/C	G/G	G/A	21	C/T	G/G	G/G		
5	C/C	G/A	G/A	22	C/T	G/G	G/G		
6	C/C	G/A	G/G	23	C/T	G/G	G/G		
7	C/C	G/G	G/G	24	C/T	G/G	G/G		
8	C/C	G/G	G/G	25	C/T	G/G	G/G		
9	C/C	G/G	G/G	26	C/T	G/A	G/A		
10	C/C	G/A	G/G	27	C/T	G/A	G/A		
11	C/C	A/A	A/A	28	C/T	G/A	G/A		
12	C/C	G/G	G/A	29	C/T	G/A	G/G		
13	C/C	G/A	G/A	30	C/T	G/A	G/A		
14	C/C	G/A	G/A	31	C/T	A/A	G/A		
15	C/C	G/A	G/A	32	T/T	G/G	G/A		
16	C/T	G/G	G/A	33	T/T	G/A	A/A		
17	C/T	G/G	G/A	34	T/T	G/A	G/G		
;									
ODS OUTPUT HAPLOTYPEFREQ = FREQOUT(KEEP = HAPLOTYPE FREQ);								调用 HAPLOTYPE 过程, 求得可能存在单体型及计算单体型概率。	
ODS HTML;									
PROC HAPLOTYPE DATA = nkkysj16_4 GENOCOL CUTOFF = 0.0075;									
VAR M1 – M3; RUN;								利用数组将 HAPLOTYPE 过程计算得到的单体型及其概率重新建立名为 HAPFREQ 的新数据集。	
DATA HAPFREQ;									
SET FREQOUT;									
ARRAY M{3} \$1;									
DOi = 1 TO 3;									
M{i} = SUBSTR(HAPLOTYPE, 2 * i – 1, 1);									
END;									
DROP HAPLOTYPEi;									
RUN;									
PROC HTSNP DATA = HAPFREQ METHOD = im CUTOFF = 0.05 SEED = 123								调用 HTSNP 过程计算得到标签 SNP。	
NOSUMM;									
VAR M1 – M3;									
FREQ FREQ;									
RUN;									
ODS HTML CLOSE;									

主要分析结果及解释如下。

The HAPLOTYPE Procedure
Haplotype Frequencies

Number	Haplotype	Freq	Standard Error	95% Confidence Limits	
1	C-A-A	0.12518	0.04043	0.04594	0.20442
2	C-A-G	0.04426	0.02513	0.00000	0.09351

3	C-G-A	0.05908	0.02880	0.00262	0.11553
4	C-G-G	0.44795	0.06075	0.32888	0.56702
5	T-A-A	0.04306	0.02480	0.00000	0.09166
6	T-A-G	0.03750	0.02321	0.00000	0.08299
7	T-G-A	0.05210	0.02715	0.00000	0.10531
8	T-G-G	0.19087	0.04801	0.09677	0.28497

The HTSNP Procedure				
htSNP Evaluation				
Rank	HTSNP1	HTSNP2	PDE	
1	M1	M3	0.9814	

由输出结果可知，共有 8 种可能存在的单体型，第 1 个和第 3 个位点反映了全部单体型的 98.14% 的信息。

专业结论：第 1 和第 3 个位点可作为 htSNPs。

16.5 一般人群病例对照遗传资料的关联分析

病例 – 对照研究是目前进行疾病关联分析应用最为广泛的设计类型。该方法是在应用家系连锁分析和连锁不平衡分析完成对多基因疾病相关基因在染色体初步定位和精细定位的基础上，通过比较候选位点等位基因型频率、等位基因频率的差异，来判断候选基因与疾病间是否存在关联，从而排除不相关的位点，得到疾病相关位点。根据选择对照的方式，可将病例对照研究设计分为一般人群对照和家系人群对照，本节主要介绍一般人群对照的关联研究。

该研究对照是从病例的源人群中随机抽样得到的，并按照混杂因子与病例进行匹配。通常使用一般 χ^2 检验、Armitage 趋势检验、多元 logistic 回归分析等统计分析方法。

【例 16-5】 观察血清脂联素 (APN) 水平及 APN 基因启动子区域单核苷酸多态性 (SNP) 频率与高血压 (HP) 及其合并症间的相关性，探讨脂联素基因 (aPM1) 多态性与 HP 及合并症间的遗传危险因素，并确立一种高效、准确、经济的检测 APN 基因实用筛查方法，供临床检测应用。脂联素基因 – 11377G/C 基因型频率与 HP 的关系见表 16-5，试分析脂联素基因 – 11377G/C 基因型是否与高血压病存在关联。

表 16-5 HP 组与对照组脂联素基因 – 11377G/C 基因型分布

组别	人数				合计
	基因型	GG	GC	CC	
对照组		2	19	37	58
HP 组		9	77	96	182
合计		11	96	133	240

例 16-5 资料的程序名为 SAS16_5.SAS。

程 序	说 明
data nkkysj16_5; input disease \$; do i = 1 to 3;	建立数据集 以下是输入变量，G1、G2 为脂联素基因 – 11377G/C 的基因型。

续表

程 序	说 明
<pre>input G1 G2 n@@ ; do j = 1 to n; output; end; end; cards; U 1 1 2 1 2 19 2 2 37 A 1 1 9 1 2 77 2 2 96 ; run; ODS HTML; PROC SORT; BY DISEASE; RUN; PROC ALLELE DATA = nkkysj16_5 PERMS = 10000 BOOT = 1000 SEED = 123 ; VAR G1 G2; BY DISEASE; RUN; PROC CASECONTROL DATA = nkkysj16_5 PREFIX = SNP; VAR G1 - G2; TRAIT DISEASE; RUN; PROC PRINT HEADING = H; RUN; ODS HTML CLOSE;</pre>	<p>U 代表对照组，A 代表病例组，使用 Allele 过程计算两组人群基因型频率是否服从哈代 - 温伯格平衡定律。</p> <p>调用 CASECONTROL 过程，对病例组、对照组与位点基因型的关联性进行χ^2 检验和趋势检验。</p>

主要分析结果及解释如下。

The ALLELE Procedure										
DISEASE = A										
Marker Summary										
Locus	Number of Indiv	Number of Alleles	Polymorph Info Content	Heterozygosity	Allelic Diversity	Chi-Square	Test for HWE			
							DF	Pr >	ChiSq	Prob Exact
M1	182	2	0.3113	0.4231	0.3857	1.7044	1	0.1917		0.2547
DISEASE = U										
Marker Summary										
Locus	Number of Indiv	Number of Alleles	Polymorph Info Content	Heterozygosity	Allelic Diversity	Chi - Square	Test for HWE			
							DF	Pr >	ChiSq	Prob Exact
M1	58	2	0.2674	0.3276	0.3179	0.0536	1	0.8170		1.0000

Obs	Locus	NumTraitA	NumTraitU	ChiSq-Genotype	ChiSqAllele	ChiSqTrend	dfGenotype	dfAllele	dfTrend	ProbGenotype	ProbAllele	ProbTrend
1	SNP1	182	58	2.18603	1.86607	2.02559	2	1	1	0.33521	0.17193	0.15467

由输出结果可以得到，两组人群的基因型概率均符合哈代 – 温伯格平衡定律。高血压疾病的发生与脂联素基因 – 11377G/C 基因型的不同是没有关联的, $\chi^2 = 2.18603$, $P = 0.33521 > 0.05$ 。当数据不满足哈代 – 温伯格平衡定律时，应使用 Armitage 检验的结果。

专业结论：脂联素基因 – 11377G/C 基因型不同与高血压不存在关联。

在对一般人群病例对照进行关联分析时，需要注意的一个问题是人群代表性问题。所谓代表性主要是指拟测量的等位基因频率与源人群中的等位基因频率相等，并且这些研究对象最好是通过普查或随机化抽样得到的。通常，获取有代表性的病例样本相对比较容易，而对照的选择就相对比较困难，因此，病例对照研究设计面临的主要挑战就是如何获取与病例相匹配的对照样本。为了确保可以获得合理的关联结论，就必须要求：所选择的对照应该具有人群代表性，应与病例来自相同的人群，对照的选择应和病例的选择同期进行等。除此之外，还可以和病例进行种族、地区、年龄和性别等因素的匹配。

16.6 家系数据的关联分析

利用家庭成员作为对照是按祖先起源匹配的最好办法，以遗传背景一致的家庭成员作为对照，可以很好地解决人群分层问题。自 20 世纪 90 年代以来，这种设计在遗传流行病学研究使用非常广泛。根据家庭成员不同，又可分为几种，分析方法也不尽相同。在 SAS/Genetics 中，对家系数据的关联分析提供了 family 过程，主要针对家系数据进行 TDT、s-TDT、SDT 检验。

16.6.1 父母亲作对照的关联分析原理介绍

由于患者及其双亲具有相似的遗传背景，有研究者提出了病例 – 父母亲对照的设计方法。该设计以患者的双亲作为对照，检验与疾病发病有关的遗传标志或者与其相邻位点存在连锁不平衡的等位基因的关系。

分析方法：传递不平衡检验 (transmission disequilibrium test, TDT)。这种检验方法只需要收集一些核心家庭数据，每个这样的家庭只有一个患病小孩和他们的父母亲。具体的抽样方法为：首先找到一个患病的个体，然后找到该个体的父母，并测定父母和患病者的标记基因型。假定标记位点仅有两个等位基因 M_1 和 M_2 。如果父亲和母亲遗传给患病者的等位基因可以完全确定，那么 n 个这样的家庭有 $2n$ 次传递 (父→子，母→子)，传递过程如表 16-6 所示。

表 16-6 n 个后代的 $2n$ 个父母传递和未传递等位基因 M_1 和 M_2 的数目

传递的等位基因	等位基因数			合计
	未传递的等位基因	M_1	M_2	
M_1		a ($P = M_1 M_1 \rightarrow M_1$)	b ($P = M_1 M_2 \rightarrow M_1$)	$a + b$
M_2		c ($P = M_1 M_2 \rightarrow M_2$)	d ($P = M_2 M_2 \rightarrow M_2$)	$c + d$
合计		$a + c$	$b + d$	$2n$

这个 2×2 表可以和配对病例对照设计的 2×2 表对应，纯合父母的传递的等位基因是没有信息的，在零假设下，基因型为杂合 ($M_1 M_2$) 的父母分别传递等位基因 M_1 和 M_2 的概率是相同的 (0.5)，如

果 M_1M_2 即为寻找的致病基因的基因型, M_1 为致病等位基因, 或者 M_1 与致病基因在同一配子上且距离致病基因非常近, 此时 b 将大于 c , 检验统计量为:

$$\chi^2 = \frac{(b - c)^2}{b + c}$$

(16-14)

由于 TDT 检验是基于同一个家庭的虚拟对照的形式进行病例对照研究的, 所以, 该方法对人群分层不敏感, 可以避免由此产生的虚假联系。此外, 采用父母亲作为对照的研究设计还可以很好地分析遗传和环境的交互作用, 且所需的样本含量比较小。因此, 病例 - 父母亲对照的设计是一种很有效的检验疾病与候选位点是否关联的分析方法。该方法已经在很多疾病, 如高血压、2 型糖尿病、精神分裂症等的研究中得到了广泛应用。

16.6.2 同胞作对照的关联分析原理介绍

通常以患者未患病的同胞作为对照, 通过比较同胞的等位基因或者基因型来检测遗传标记是否与疾病位点存在关联。分析方法通常使用以下几种。

1. 同胞-传递不平衡检验(sib-transmission disequilibrium test, s-TDT)

通过比较患病同胞的等位基因频率和没有关联时的期望值之间的差别来检验遗传标记与疾病是否存在关联。但要求每组同胞必须满足下面的条件: ① 每个家庭同胞中至少有一个患者和一个未患病的同胞; ② 每个家庭同胞中所有成员的基因型不能完全相同。将数据整理成如表 16-7 所示的形式。

表 16-7 s-TDT 分析的数据结构

家系号 (i)	患病状态	同胞数			合计	
		基因型	1/1	1/2		2/2
i	患病		$n_{1\ i}$	$n_{2\ i}$	—	a_i
	未患病		n_{3i}	$n_{4\ i}$	—	u_i
	合计		r_i	s_i	—	t_i

s-TDT 检验的统计量为:

$$Z = (Y - A) / \sqrt{\text{Var}} \text{ 或 } Z' = (|Y - A| - 1/2) / \sqrt{\text{Var}}$$

(16-15)

Z 近似服从正态分布。式中,

$$Y = \sum (n_{1i} + n_{2i}) = \sum a_i$$

(16-16)

$$A = \sum (2r_i + s_i) a_i / t_i$$

(16-17)

$$\text{Var} = \sum au[4r(t - r - s) + s(t - s)]/[t^2(t - 1)]$$

(16-18)

a 表示患者数目, u 表示未患病同胞数目, 家系中所有同胞的数目为 $t = a + u$, r 表示基因型为 1/1 的数目, s 表示基因型为 1/2 的数目。

2. 同胞组不平衡检验(sibship disequilibrium test, SDT)

在每一个同胞对中, m_A^1 和 m_U^1 分别表示每个同胞对中患者和未发病同胞携带变异等位基因 1 的均数, 即

$$m_A^1 = \{ \text{患者携带等位基因 1 的个数} \} / \text{同胞对中发病者人数};$$

$$m_U^1 = \{ \text{对照携带等位基因 1 的个数} \} / \text{同胞对中未发病人数};$$

$d^1 = m_A^1 - m_U^1$ 表示患者和对照携带变异等位基因 1 的均数之差。当 $d^1 = 0$ 时, 这样的同胞对资料被删除; 用 b 表示 $d^1 > 0$ 的同胞对数目; c 表示 $d^1 < 0$ 的同胞对数目。

SDT 检验的统计量为：

$$T = \frac{(b - c)^2}{b + c}$$

(16-19)

近似服从自由度为 1 的 χ^2 分布。

3. 同胞对关联检验法 (sib association test, SAT)

若患者有多个发病同胞，则选择基因型与患者有最大差别的一个同胞作为对照。每个同胞对中，包括了一个患者和一个患者未发病的同胞。考虑双等位基因时，双亲和患者基因型有如表 16-8 所示的三种有效的组合。

表 16-8 同胞对关联检验法数据结构

双亲基因型	患者基因型	患者基因型频率	对照基因型	对照基因型概率	得分计数
AA, AB	AA	0.5	AB	$1 - 0.5^S$	t_{AB}
	AB	0.5	AA	$1 - 0.5^S$	t_{BA}
BB, AB	BB	0.5	AB	$1 - 0.5^S$	t_{BA}
	AB	0.5	BB	$1 - 0.5^S$	t_{AB}
	AA	0.25	BB	$1 - 0.75^S$	$2t_{AB}$
AB, AB	AB	0.5	AB	$0.75^S - 0.25^S$	t_{AB}
			AA	$0.5(1 - 0.5^S)$	t_{BA}
	BB	0.25	BB	$0.5(1 - 0.5^S)$	t_{AB}
			AA	$1 - 0.75^S$	$2t_{BA}$
			AB	$0.75^S - 0.25^S$	t_{BA}

其中，S：作为对照的同胞数目； t_{AB} ：患者双亲传递等位基因 A 给患者、传递等位基因 B 给对照； t_{BA} ：患者双亲传递等位基因 B 给患者、传递等位基因 A 给对照。

检验统计量：

$$Z = (T_{AB} - N_1/2 - N_2)/\text{sqrt}(N_1/4 + N_2)$$

(16-20)

该统计量近似服从正态分布。其中 T_{AB} 表示标记为 t_{AB} 的次数， N_1 表示标记为 t_{AB} 或者 t_{BA} 的次数， N_2 表示标记为 $2t_{AB}$ 或 $2t_{BA}$ 的次数。

除上述方法之外，分析方法还有很多，如表型不一致同胞对分析 (discordant sib pair analysis, DSP)、同胞对内重复抽样方法 (within sibship paired resampling method, WSPR) 等。在选取对照方面，除选取一般同胞作为对照外，有时也可选择堂 (表) 兄妹同胞作为对照并且选择相应的分析方法。限于篇幅所限，请读者参阅其他相关文献。

【例 16-6】 某研究探讨 β_2 肾上腺素能受体基因 (β_2AR) 的多肽性与中国人哮喘易感性的连锁关系及对疾病调节作用的重要性。用 PCR/RFLP 法检测 β_2AR 基因编码区第 16 位多态性位点及 IL-4 基因启动子区-589 位多态性位点 (2 个 SNP 位点)，对 15 个家系共 69 份数据进行分析，数据如表 16-9 所示。试分析哮喘病是否同两个位点存在关联。

表 16-9 哮喘家系样本 ADR 基因第 16 位, IL-4 基因启动子区-589 位多态性位点资料

家系编号	个体编号	父亲编号	母亲编号	性别 (M/F)	疾病 (1/2)	A16G	589
1	1	0	0	M	2	A G	T T
1	2	0	0	F	1	A A	T C
1	101	1	2	M	2	A G	T T
1	102	1	2	M	2	A G	T C
2	3	0	0	M	1	A A	T C
2	4	0	0	F	1	A G	C C

续表							
家系编号	个体编号	父亲编号	母亲编号	性别 (M/F)	疾病 (1/2)	A16G	589
2	103	3	4	F	2	A A	C C
2	104	3	4	M	2	A A	T C
2	105	3	4	M	2	A A	C C
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
15	29	0	0	M	1	A G	T T
15	30	0	0	F	1	G G	T T
15	138	29	30	M	2	G G	T T
15	139	29	30	M	2	G G	T T

注：疾病：2 代表发病,1 代表正常,在实际计算中应增大样本含量，即应再多收集家系资料。

【例 16-7】 某研究机构研究某种复杂的遗传疾病，对 10 个家系进行分析，检测个体两个基因的基因型，数据如表 16-10 所示，试分析该遗传疾病与两个基因是否存在关联，进行 TDT、s-TDT、SDT 检验。

表 16-10 10 个家系样本 2 个基因基因型资料

家系编号	个体编号	父亲	母亲	性别	disease	基因 1 的基因型		基因 2 的基因型	
1	1	0	0	1	1	7	8	7	2
1	2	0	0	2	1
1	401	1	2	1	1	7	2	7	6
1	402	1	2	1	1	8	2	7	6
1	403	1	2	1	1	7	2	2	7
1	404	1	2	2	2	8	2	7	7
2	3	0	0	1	1	4	4	1	3
2	4	0	0	2	1
2	405	3	4	2	1	4	6	1	7
2	406	3	4	2	2	4	4	3	7
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10	19	0	0	1	1	6	4	2	7
10	20	0	0	2	1
10	427	19	20	2	2	4	4	7	2

SAS 程序中重要内容的说明如下。
例 16-6 资料的程序名为 SAS16_6. SAS。

程 序	说 明
<pre>DATA nkkysj16_6; INPUT PED INDIV FATHER MOTHER SEX \$ DISEASE (A1 – A4) (\$); CARDS; /* 录入表 16-9 中的数据 */ ; RUN; ODS HTML; PROC FAMILY DATA = nkkysj16_6 PREFIX = SNP OUTSTAT = AAA TDT; ID PED INDIV FATHER MOTHER; TRAIT DISEASE/AFFECTED = 2; VAR A1 – A4; RUN; PROC PRINT DATA = AAA; RUN; ODS HTML CLOSE;</pre>	<p>建立数据集</p> <p>以下是输入变量，A1、A2 为位点 1 的基因型，A3、A4 为位点 2 的基因型。</p> <p>调用 family 过程对两个位点进行 TDT 检验。</p>

例 16-7 资料的程序名为 SAS16_7. SAS。

程 序	说 明
DATA nkkysj16_7; INPUT PED ID F_ID M_ID SEX DISEASE A1 – A4; CARDS; /* 录入表 16-10 中的数据 */ ; RUN; ODS HTML; PROC FAMILY DATA = nkkysj16_7 PREFIX = MARKER TDT SDT STDT; ID PED ID F_ID M_ID; VAR A1 – A4; TRAIT DISEASE / AFFECTED =2; RUN; PROC PRINT; RUN; ODS HTML CLOSE;	建立数据集 以下是输入变量，A1、A2 为基因 1 的基因型，A3、A4 为基因 2 的基因型。 调用 family 过程对两个基因进行 TDT、sTDT、SDT 检验。

主要分析结果及解释

	Obs	Locus	ChiSqTDT	dTDT	ProbTDT
	1	snp1	0.44444	1	0.50499
	2	snp2	1.33333	1	0.24821

这是例 16-6 的输出结果，结果表明，对于位点 1, $\chi^2 = 0.44444$, $P = 0.50499 > 0.05$ ；对于位点 2, $\chi^2 = 1.33333$, $P = 0.24821 > 0.05$ ，疾病与两个位点均无关联，在本例调查的家系中，大多数家系的孩子都是患病的，所以不能进行 s-TDT、SDT 检验。

专业结论：β₂AR 基因编码区第 16 位多态性位点及 IL-4 基因启动子区-589 位多态性位点基因型频率差别与哮喘病之间没有关联。

Obs	Locus	ChiSqTDT	ChiSqSTDT	ChiSqSDT	dTDT	dTSTDT	dTSDT	ProbTDT	ProbSTDT	ProbSDT
1	Marker1	4.80000	7.07688	4.14286	4	6	7	0.30844	0.31380	0.76318
2	Marker2	1.33333	6.38259	7.00000	2	5	6	0.51342	0.27075	0.32085

这是例 16-7 的输出结果，结果表明，经过 TDT 检验、s-TDT 检验、SDT 检验，该遗传疾病与两个位点均无关联。

专业结论：所研究的两个位点基因型频率的差别与遗传病之间没有关联。

16.7 本章小结

本章主要介绍了 SAS/Genetics 模块中 allele、haplotype、htsnp、casecontrol、family 这 5 个过程计算基因频率、验证哈代 – 温伯格平衡定律、尝试性进行连锁不平衡定位、寻找 htSNPs、对一般病例对照数据和家系数据进行关联分析，并对每类问题给出相应的理论知识，以及结合实例利用 SAS 进行了详细的分析，对结果进行了阐释。对于某些遗传统计分析方法，如连锁分析等，需要结合 SAS/IML 模块、SAS/STAT 模块进行计算，本章并未涉及，请读者参照其他章节程序和有关文献自行计算。

(郭 晋)

第 17 章 提高医学论文统计学质量的策略

一般来说,科研人员在撰写科研论文中都需要运用统计学。事实上,不应当等到研究工作(调查和试验等)结束并开始着手撰写论文时,才想起要运用统计学以提升学术论文的“档次”。正确的做法是:在科研工作开始时,就应当充分发挥统计学的作用,并且,应自始至终正确运用统计学。具体表现是:制订出科学完善的课题设计方案、在课题实施过程中实时进行严格的质量控制、坚持高质量的数据记录和数据管理、依据分析目的和多方面情况选用合适的统计分析方法处理数据、按各杂志的要求将统计学正确地应用和表达于论文之中。

17.1 设计阶段存在的问题

科研设计阶段对任何研究来说都是最重要的阶段。在此阶段发生的错误、缺点和不足对研究结果的有效性和可靠性有巨大的负面影响,因为它们会影响所有后续研究的各个阶段的质量和进程,所以该阶段的错误不容忽视。

首先,当我们规划一个研究项目时,制定研究目的是非常重要的,主要、次要的疗效指标和研究终点的设定应该在最初的研究计划中明确下来,研究方案要经过反复商讨。研究中需要做的统计分析以及科学的假设应该预先明确指出,特别是当研究中需要做多个假设检验时,这点尤为重要。对于高品质的统计工作,关键是要在规划阶段预先考虑一个先验的效应值、选择恰当的检验效能,以便估算出合适的样本量,以确保你的研究有足够的统计检验效能,以检测出所观察的处理效果,并且尽可能避免Ⅱ型错误。此外,所使用的样本大小在任何阶段应始终记录在研究方案和计划书中。还要尽可能地采用随机样本,在抽取和分配受试对象时,采取恰当的随机化和盲法以避免、至少大幅度降低潜在的偏倚。

然而,我们必须十分清楚一点,那就是所有的统计推断只有对随机样本才是有效的,没有必要坚持说旧的数据如何收集,只有随机化才是可靠的,因此随机化抽样应该在撰写科研设计方案的初稿中强制被写入并提供具体实现的途径。

如果使用对照组,应确保研究组在基线时的相似性,即具有均衡可比性,以保证消除组间的不同质性。如果组间在基线状态不具备可比性,那么干预后的差异就有可能受基线的影响,同样不具备可比性,研究的结果没有实际意义。然而,当重要的非试验因素的影响在对比组之间出现了严重的偏差,即使应用统计学的假设检验测试基线的平衡既无必要也不可取,也很难通过统计分析技术(如协方差分析)彻底消除其影响;即使对比组间在基线上没有统计学差异,也不能因此断定组间是等价的,在研究样本量较小或遗漏了某些重要非试验因素时尤其如此。因为样本量较小时会缺乏检验效能,而遗漏了重要非试验因素就好比“抓了芝麻丢了西瓜”。我们只能估计在无效假设下获得该数据的概率,而没有其他方式。其实我们真正需要的,而且也是最重要的,是找准找全对研究目的和主要评价指标有影响的重要非试验因素,并在设计阶段将其纳入严格控制之下,在试验或调查实施阶段实时对它们进行有效监控,作为辅助或补充控制措施,再采取合适的统计分析技术对其进行校正。

与研究的计划和设计阶段相关的错误主要表现在以下几个方面:

- (1) 研究目标和主要结局变量不明确或没有在方案中清楚地表明。
- (2) 设计方案有严重的“硬伤”。
- (3) 未报告样本量的大小。
- (4) 没有报告研究中中途终止的病例。
- (5) 没有根据效应估计值和检验效能大小计算样本量。
- (6) 没有对研究的无效假设做明确的陈述。
- (7) 未采用随机化方法或者对采用的随机化方法未明确陈述。
- (8) 需要采用盲法时未采用。
- (9) 未陈述基线特征及其基线时对比组间的可比性。
- (10) 缺乏合理的对照组或犯了对照不全的错误或采用了不适当的对照组。
- (11) 基线特征相等的检验方法不正确。

样例设计阶段的问题展示及对差错的辨析与释疑

【例 17-1】 为探讨糖尿病(DM)大鼠肺组织糖基化终末产物(AGEs)、结缔组织生长因子(CTGF)的变化及其相关性。将 48 只 3 月龄体重为 185 ~ 210g 雄性 SD 大鼠(购于天津药物研究院)随机分入 DM 组和正常对照(NC)组各 24 只。两组分别以 60mg/kg STZ 和等量生理盐水尾静脉注射, 3 天后测 FBG > 16.65mmol/L 为 DM 模型成立。分别在试验第 4、12、20 周末处死两组大鼠各 8 只, 即取肺脏, 10% 中性甲醛固定, 常规脱水, 石蜡包埋, 测量两组的 AGEs 和 CTGF, 结果为(1)AGEs 测量结果: 第 4 周 DM 组大鼠肺泡上皮细胞以及支气管黏膜上皮细胞、血管内皮细胞及平滑肌细胞即可见大量 AGEs 阳性细胞, 其平均灰度值低于 NC 组($P < 0.05$), 并随病程进展而逐渐降低($P < 0.01$)(表 17-1); (2)CTGF 测量结果: 4 周 DM 组大鼠支气管黏膜上皮细胞、血管内皮细胞、肺间质细胞可见少量 CTGF 阳性细胞, 12、20 周 DM 组大鼠可见大量阳性细胞, 其平均灰度值低于 NC 组($P < 0.01$), 并随病程进展逐渐降低($P < 0.01$)(表 17-2); (3)AGEs、CTGF 的 Pearson 相关分析: DM 大鼠肺组织中 AGEs 和 CTGF 阳性细胞的平均灰度值之间呈正相关($r = 0.939$; $P < 0.01$)。

表 17-1 两组大鼠肺组织 AGEs 平均灰度值比较($\bar{x} \pm s$)

组别	鼠数(只)	4 周	12 周	20 周
NC	8	104.75 ± 8.20	93.50 ± 7.56	142.38 ± 19.76
DM	8	93.92 ± 7.92 * △	76.25 ± 6.76#△	43.63 ± 7.96#△

注: 与对照组相比, * 代表 $P < 0.05$, #代表 $P < 0.01$; DM 组 4、12、20 周比较, △代表 $P < 0.01$ 。

表 17-2 两组大鼠肺组织 CTGF 平均灰度值比较($\bar{x} \pm s$)

组别	鼠数(只)	4 周	12 周	20 周
NC	8	94.00 ± 7.69	92.63 ± 9.12	97.75 ± 6.27
DM	8	93.00 ± 4.34#	75.37 ± 6.05 * #	46.13 ± 4.64 * #

注: 与对照组相比, * 代表 $P < 0.01$; 4、12、20 周 DM 组内比较, #代表 $P < 0.01$ 。

【分析与解答】 本例为内科学的动物试验, 实际上, 原作者的研究目的有三个: 一是看处理因素的作用, 在此可看成药物的作用, 有两个水平, 即 60mg/kg STZ 和生理盐水; 二是看时间因素的作用, 有三个水平, 即 4 周、12 周和 20 周; 三是看两个指标之间的联系, 即 AGEs 和 CTGF 之间的相关性。本例的问题在哪儿呢? 首先我们来看, 本例实际上是两个因素的试验研究问题。因为各个因素的水平数不相等, 当缺乏专业知识不能知晓两个因素对结果的影响有无主次之分时, 此试验设计实际应为 2×3 的析因设计, 而原作者却将其视为单因素多水平设计, 由于设计类型的判定错

误，导致后面的一系列做法和结果均出现错误，原作者在将 NC 与 DM 展开比较的同时，又在 NC 和 DM 各自的组内进行了比较，这样不但与研究的目的不符合，而且会增大犯 I 型错误的概率。另外，本文的原作者还对 DM 鼠的 AGEs 和 CTGF 之间的相关性展开了研究，得出的结论是：DM 大鼠肺组织中 AGEs 和 CTGF 阳性细胞的平均灰度值之间呈正相关($r=0.939$; $P<0.01$)。从原作者给出的有关曲线来看，采用的是 24 只老鼠，但由于 24 只鼠的 AGEs 和 CTGF 数值都是在不同的时间点上测量得到的，放在一起并不具备可比性，也无实际意义。也就是说，要研究两个指标的相关性时，两个指标必须在同一个时间点测量。鉴于上述问题，本课题得出的结果也就没有多少实际意义了。

本文的正确做法应该是采用 2×3 的析因设计类型，具体模型的形式如下：

对所收集的数据在保证正态分布、方差齐性的条件下，采用析因设计的方差分析，分别看两个因素对两个指标的影响，以及两个因素之间是否具有交互作用。如果两个因素之间无交互作用，且不同时间点的各个指标间存在统计学差异，则进行各个时间点之间的比较；若交互作用有统计学意义，应在控制其中一个因素条件下，对另一个因素全部水平进行两两比较。若两个定量指标在专业上有联系，希望对其同时分析，还可采用两因素析因设计二元定量资料方差分析。这才是正确的方法。

对于 AGEs 和 CTGF 两个指标之间的相关性研究，只能在同一个小组内同一个时间点测量两个指标(共 8 对数据)，然后研究其相关性。因为 4 周的 AGEs 和 12 周的 CTGF 之间并不具备进行相关分析的实际意义。

【例 17-2】 本研究旨在观察 OHA(口服降糖药)治疗失效的糖尿病患者采用不同疗程 CSII(连续胰岛素皮下注射)后胰岛 β 细胞功能改善的情况。2003 年 3 月—2004 年 3 月在某院内内分泌科住院的 T2DM 患者(依据 1999 年 WHO 诊断标准)，经门诊 3 个月指导饮食和运动治疗，并联合使用磺脲类降糖药、二甲双胍和(或)阿卡波糖治疗后，血糖仍控制不好的共 48 例，按 CSII 疗程将入选者分为 7 天(7d)组 10 例、14 天(14d)组 18 例和 28 天(28d)组 20 例，三组基础病变、性别、年龄和糖尿病史相匹配($P>0.05$)，均计算 CSII 治疗第 1 天和最后一天血糖达标时胰岛素的基础用量和餐前日追加量。28d 组 1 年后复查 PG 和 HbA_{1c}，统计恢复使用 OHA 的比例。数据以 $\bar{x} \pm s$ 表示， t 检验比较治疗前后的差别，方差检验比较三组参数。结果为：(1)CSII 治疗后的降糖效果：CSII 治疗后 48 例均迅速控制血糖，并在停泵后确保血糖控制在较为理想状态，HOMA- β 明显提高，而 HOMA-IR 明显下降($P<0.01$)(表 17-3)。(2)CSII 治疗后胰岛素日用量：CSII 治疗第 1 天血糖达标时胰岛素的基础用量和餐前日追加量三组无统计学差异($P>0.05$)。14d 和 28d 组疗程结束血糖达标时胰岛素的基础用量和餐前日追加量均较治疗第 1 天减少($P<0.05$, $P<0.01$)。28d 组 CSII 平均胰岛素日用总量明显低于 7d 组($P<0.05$)(表 17-4)。

表 17-3 T2DM 患者 CSII 治疗前后 PG、HOMA- β 、HOMA-IR 的变化及血糖控制所需时间比较($\bar{x} \pm s$)

CSII	FPG (mmol/L)	2hPG (mmol/L)	HOMA- β	HOMA-IR	达标时间 (d)
之前	15.0 \pm 4.0	22.0 \pm 6.0	45 \pm 20	9.2 \pm 3.5	5 \pm 2
之后	5.0 \pm 1.5 *	8.0 \pm 1.3 *	112 \pm 60 *	5.3 \pm 1.4 *	5 \pm 2

注：与治疗前比较，* 代表 $P<0.01$ 。

表 17-4 三组 CSII 各时期血糖达标时的胰岛素用量比较($\bar{x} \pm s$)

组别	第 1 天剂量		第 7、14、28 天剂量		平均胰岛素日用总量
	基础量	餐前追加量	基础量	餐前追加量	
7d	0.8 \pm 0.20	28 \pm 6	0.7 \pm 0.18	28 \pm 4	0.8 \pm 0.1
14d	0.8 \pm 0.18	27 \pm 6	0.6 \pm 0.12 *	25 \pm 4 *	0.7 \pm 0.1
28d	0.8 \pm 0.20	28 \pm 6	0.5 \pm 0.15#	20 \pm 3#	0.5 \pm 0.2

注：CSII 血糖达标与 CSII 结束时比较，#代表 $P<0.05$ ，* 代表 $P<0.01$ 。

【分析与解答】 本例为内科学的临床试验，目的在于判断不同疗程 CSII 后胰岛 β 细胞功能改善的情况。临床试验应该遵循科研设计中的随机、对照、重复和均衡等原则。本研究中的第一个错误是未采用随机化原则，只是提到将患者按照不同疗程分为 7 天组、14 天组和 28 天组，没有提到随机化的问题。这就难免会有混杂因素，虽然研究者提到三组基础病变、性别、年龄和糖尿病史相匹配 ($P>0.05$)，但是影响糖尿病的其他因素却不一定相同，比如生活习惯、心态等对糖尿病病人的血糖有较大影响。正确的做法不是按照疗程分组，而是将病人随机化分为 3 个组后，分别采用 7 天、14 天和 28 天 3 个疗程。第二个错误就是对照的问题，原作者的目的是观察 OHA 治疗失效的糖尿病患者采用不同疗程 CSII 后胰岛 β 细胞功能改善的情况，按照设计方案，应该是 3 个治疗组互相对照，如果做前后自身对照，也应该是对各个组的前后指标进行对照，而不是将 3 个组合并后对照，因为 3 个组的治疗时间不同，合在一起不具备可比性，研究者采用将 3 个治疗组合并做前后比较，并做前后配对统计分析，这是不合适的。第三个错误是研究者未采用盲法，盲法也是临床试验应该采取的一个消除或降低心理因素对结果造成影响和提高受试者依从性的重要举措。因为病人住院在同一个科室，如果盲法做不好，很容易产生偏倚。第四个错误是在统计分析方法选择方面，采用的是单变量的还是多变量的方差分析并不是十分清楚，因为研究者并没有给出具体的统计量，只给出了 P 值。第五个错误是在结果的表达方面。在表 17-4 中，原作者可能表达了 3 种含义：一是 3 个疗程组之间第 1 天剂量的比较（包括基础量和餐前追加量），这里采用的应该是多变量方差分析，但是原作者没有具体说明。二是将第 7、14 和 28 天的剂量与第 1 天的剂量做了比较，采用的应该是配对 t 检验，原作者没有说明；三是将 3 个疗程的最后一天剂量做了比较，采用的应该也是方差分析，原作者也没有说明方差分析的结果及两两比较的结果分别是什么。另外，第三个纵标目的表达有问题，第 7、14 和 28 天的剂量，实际就是 3 个不同疗程中最后一天的剂量。应该改为“最后一天剂量”。写成“第 7、14、28 天的剂量”，让读者误解为将不同疗程剂量合并了。总之，表 17-4 表达的内容繁杂，模糊不清，应该分为两张或三张表分别表达，并对纵表目做修改。

17.2 数据分析阶段存在的问题

当进行数据的统计分析和运用统计假设检验或参数估计技巧时，应十分清楚：每一种方法都是基于几个基本假设的，这些假设至少基本成立才能获得正确的、有意义的结果。然而，即使是简单的假设检验方法，如通用的 t 检验或卡方检验也经常会滥用，因为假设检验的条件在检验使用前没有做过充分的评价。此外，当采用 t 检验或卡方检验时，要注意选择正确的形式。比如 t 检验，就包括了单因素两水平设计一元定量资料 t 检验、单组设计一元定量资料（与一个标准值比较） t 检验和配对设计一元定量资料 t 检验。它们各自具有不同的适用条件。再比如，四格表的卡方检验也是有前提条件的，如果四个格子中有一个格子内的理论数小于 5，卡方检验的条件就不满足了。因为在这种情况下，近似超几何分布不再可靠。如果样本量很小，应该采用 Yates 连续性校正或者采用 Fisher 的确切概率法，后者更准确一些。此外，在种类繁多的统计分析方法中选择最适合和最强大的一个，不但不是微不足道的，反而是异常重要的。因为要考虑非常多的细节。

如果研究中必须采用多个统计检验，则需要注意假阳性率及其错误发生率的潜在膨胀，此时往往需要对多重比较进行校正。尤其重要的是要认识到两个以上的组间比较时，要求使用参数的方差分析或非参数的相应分析方法，如果重复进行两组间的比较，假阳性结果的风险就会增大很多。然而在研究设计不佳的研究中，却经常发生做多个两组间比较的情形。其实，如果在设计之前咨询一下统计学家，可以很容易地避免这类错误。

没有在设计方案中预先指定的检验,后来拆分出许多亚组进行分析,这种做法也应避免,因为它会导致本不应出现的大量有统计学意义的结果。此外,如果多组间存在混杂因素或者在基线时组间不平衡,则必须采用多变量分析方法。下面总结统计分析中的主要错误和不足之处。

17.2.1 常被误用的统计检验方法

- (1) 采用了与数据类型不匹配的统计检验方法。
- (2) 非配对数据采用配对检验或者配对数据采用了非配对检验。
- (3) 参数检验方法的使用不正确。
- (4) 采用的方法与所做的假设不一致。
- (5) 出现了 I 型错误膨胀的情况。
- (6) 没有对做多重比较结果进行校正。
- (7) 做了不合适的事后亚组分析。
- (8) t 检验中未明确交代设计类型。
- (9) 未能证明假设的条件是否成立。
- (10) 两样本大小不相等却采用了配对 t 检验。
- (11) 对多个组采用了两组间成对比较的 t 检验方法。
- (12) 对于配对数据使用了非配对 t 检验或者对非配对数据采用了配对 t 检验。

17.2.2 使用卡方检验人们常犯的典型错误

- (1) 对于小样本没有做 Yates 连续性校正。
- (2) 在四格表资料中有一格理论频数小于 5 时仍然采用一般卡方检验。
- (3) 没有明确的声明无效假设。
- (4) 将高维列联表简单压缩成二维列联表进行一般卡方检验。
- (5) 用卡方检验分析结果变量为有序变量的二维列联表资料。

17.2.3 数据分析阶段有问题的样例展示及对差错的辨析与释疑

【例 17-3】 为探讨 TNF- α 诱导的胰岛素抵抗 (IR) 小鼠胰岛素敏感性及糖脂代谢的变化,将 23 只健康雄性 C57BL/6J 小鼠随机分为 4 组:高剂量 (H) 组、中剂量 (M) 组、低剂量 (L) 组分别给予腹腔注射 6、3、1tg/kg/d 的 TNF- α , 正常对照 (NC) 组注射等体积的生理盐水,共 7 天。采用静脉葡萄糖耐量试验 (IVGTT) 和 $3-[^3\text{H}]$ 葡萄糖为示踪剂的扩展胰岛素钳夹技术,评价小鼠胰岛素敏感性和糖脂代谢的变化。各项指标均以 $\bar{x} \pm s$ 表示,用 SPSS11.5 软件进行统计分析。组间比较经方差齐性检验后采用独立样本 t 检验,组内比较用配对 t 检验。结果如表 17-5 所示。

结果表明, TNF- α 处理后,小鼠 FBG、血浆胰岛素 (Ins) 和 FFA 水平均增高,且 H 组明显高于 NC、M 和 L 组。IVGTT 结果显示 H 组糖耐量减低,Ins 释放水平明显高于其他组。在胰岛素钳夹术中,H 组基础葡萄糖清除率 (GDR) 和肝糖输出率 (HGP) 明显高于 NC 组 ($P < 0.01$)。在钳夹稳态时,H 组血浆 Ins 水平明显高于 NC 组 ($P < 0.01$),Ins 对 FFA 的抑制作用较 NC 组明显降低 ($P < 0.01$),H 组葡萄糖输注率 (GIR) 明显低于 NC 组 ($P < 0.01$);钳夹稳态时,小鼠 GDR 明显增加,但 H 组 GDR 的增加仍明显低于 NC 组 ($P < 0.01$);钳夹结束时,NC 组 HGP 被完全抑制,而 H 组仅被抑制 59%。结论:高剂量 TNF- α (6tg/kg/d) 处理可导致小鼠糖脂代谢紊乱以及肝和外周组织的 IR。

表 17-5 四组临床和生化指标比较 ($\bar{x} \pm s$)

组别	鼠数	体重(g)	FBG (mmol/L)	Flns (mU/L)	FFA (mmol/L)	TG (mmol/L)	TC (mmol/L)	HOMA-IR
NC	6	28.3 ± 1.5	5.58 ± 0.33	12.4 ± 1.12	1.54 ± 0.11	0.57 ± 0.5	3.65 ± 0.30	3.07 ± 1.1
H	7	28.8 ± 1.9	7.86 ± 0.40 #△	32.4 ± 1.9#☆	1.78 ± 0.05 * △	0.62 ± 0.04	3.85 ± 0.32	11.8 ± 2.3#☆▲
M	5	26.9 ± 1.2	6.56 ± 0.09	19.6 ± 1.8	1.63 ± 0.05	0.58 ± 0.02	3.72 ± 0.13	5.62 ± 0.38
L	5	29.3 ± 1.0	6.38 ± 0.25	13.5 ± 1.3	1.58 ± #0.10	0.55 ± 0.04	3.64 ± 0.17	3.83 ± 0.35

注：与 NL 组比较，* 代表 $P < 0.05$ ，#代表 $P < 0.01$ ；与 L 组比较，△代表 $P < 0.05$ ；☆代表 $P < 0.01$ ；与 M 组比较，▲代表 $P < 0.01$ 。

【分析与解答】 本例为内科学的动物试验，设计遵循了随机对照重复的原则，将受试对象随机地分成了 4 个处理组，应该说，随机化及对照的选择都符合动物实验要求。但是在对指标进行统计分析时，却出现了错误，误将多组间比较的方差分析用成了 t 检验， t 检验仅适用于 2 个处理组的情形，在处理组数多于 2 个时，如果数据服从正态分布、方差齐性，可以采用相应设计一元定量资料方差分析，即 F 检验，在得出总体有统计学意义的结论后，如需具体看哪几个组之间有差别，再采用 SNK- q 检验、Dunnett- t 检验等做组间的两两比较。本研究直接用 t 检验在不同组之间进行比较，会增大犯 I 型错误的概率，易于得出假阳性的结论。

【样例 17-4】 240 例某院住院的 T2DM 患者，符合 1999 年 WHO 糖尿病诊断标准，且无急性并发症，均经口服药物或胰岛素治疗血糖控制不佳。随机分为诺和锐 30 组 124 例(男 68，女 56)，年龄 60 ± 4 岁，病程 6.3 ± 2.8 年，BMI $24.7 \pm 1.8\text{kg/m}^2$ ，FBG $10.58 \pm 3.12\text{mmol/L}$ ，2hBG $13.69 \pm 4.15\text{mmol/L}$ ，HbA1c $8.8\% \pm 0.6\%$ ；诺和灵组 116 例(男 58，女 58)，年龄 58 ± 4 岁，病程 5.9 ± 2.5 年，BMI $25.2 \pm 2.1\text{kg/m}$ ，FBG $10.26 \pm 2.59\text{mmol/L}$ ，2hBG $14.06 \pm 3.52\text{mmol/L}$ ，HbA C $8.7\% \pm 0.8\%$ 。两组差异均无统计学意义($P > 0.05$)，具有可比性。诺和锐 30 组(124 例)日 3 次诺和锐 30 皮下注射；诺和灵组(116 例)日 4 次重组人胰岛素治疗(三餐前半小时注射诺和灵，睡前注射诺和灵)。观察两组治疗前后 FBG 和 2hBG 的变化，记录胰岛素日用量及低血糖发生情况。统计学方法：两组数据用 $\bar{x} \pm s$ 表示，组间数据比较用 t 检验。结果：(1)治疗前两组 FBG、2hBG 水平均无统计学差异($P > 0.05$)；治疗后均较治疗前明显下降($P < 0.01$)，但两组间比较差异无统计学意义($P > 0.05$)。(2)两组血糖达标时间及胰岛素用量比较(表 17-6)诺和锐 30 组血糖达标所需时间较诺和灵组明显缩短($P < 0.01$)，且胰岛素用量明显减少。

表 17-6 两组病人治疗前后血糖比较 ($\bar{x} \pm s$)

组别	例数	FBG (mmol/L)		2hBG (mmol/L)		血糖达标 BG	
		基础量	餐前追加量	基础量	餐前追加量	时间	Ins 剂量
诺和锐 30	124	14.58 ± 3.12 *	5.8 ± 0.8	21.69 ± 4.15 *	7.4 ± 1.1	6.5 ± 1.1	45.5 ± 14.6
诺和灵	116	13.26 ± 2.59 *	6.2 ± 0.5	20.06 ± 3.52 *	7.8 ± 1.4	9.1 ± 1.6#	54.1 ± 11.9

注：与治疗前比较，* 代表 $P < 0.01$ ；与诺和锐 30 组比较，#代表 $P < 0.01$ 。

【分析与解答】 本例为内科学的临床试验，设计遵循了随机对照重复的原则，将受试对象随机地分成了两个处理组。在方案中也给出了统计分析的方法。但是只是笼统地说采用 t 检验，并未说明采用何种类型的 t 检验，其实本研究在对用药前后展开自身对照时采用的是配对 t 检验，而在对诺和锐 30 与诺和灵两组进行比较时采用的是两独立样本的 t 检验，二者是不同的，应该在统计分析方法中予以说明，即具体地说明对于哪个数据集采用了哪种统计分析方法。另外，两组血糖的达标时间是否服从正态分布，未做说明。如该指标不服从正态分布，则应采用非参数检验的方法。

【例 17-5】 为了探讨大黄酸改善糖尿病大鼠脂肪组织胰岛素敏感性及降血糖的作用机制。将雄性 Wistar 大鼠随机分为对照组 (NC, $n = 15$) 及糖尿病造模组 (DM, $n = 40$)，DM 组成模后随机分为糖尿病模型组 (DM-C, $n = 15$) 和糖尿病大黄酸治疗组 (DM-T, $n = 15$)。DM-T 组予大黄酸 100mg - kg · d 灌胃 11 周。试验末检测各组大鼠 FBG、Fins、TG、TC、胰岛素敏感指数 (ISI)、过氧化物酶体增殖物激活受体 (PPAR- γ) 及葡萄糖转运蛋白 4 (GluT4) 在脂肪组织的表达水平。统计分析方法：以 SPSS 10.0 软件进行分析，正态分布数值以 $\bar{x} \pm s$ 表示，两组间比较采用 t 检验或 z 检验；非正态分布测定值以中位数及四分位数表示，两组间的比较采用秩和检验。 $P < 0.05$ 为差异有统计学意义。结果如表 17-7 所示，由表 17-7 可知，17 周末 DM-C 组较 NC 组 FBG、TG 明显升高，ISI 明显降低，脂肪组织 PPAR-7 及 GluT4 蛋白表达明显降低；DM-T 组较 DM-C 组 FBG 明显降低；ISI 明显升高；脂肪组织 PPAR-蛋白表达明显升高，积分光密度有统计学差异；GluT4 蛋白表达明显升高，积分密度 (ID) 有统计学差异。

表 17-7 17 周末各组大鼠相关生化指标的比较 ($\bar{x} \pm s$)

组别	鼠数	FBG (mmol/L)	Flns (mU/L)	ISI	IS	HbA1c	TG (mmol/L)	TC (mmol/L)
NC	12	7.45 \pm 1.81	0.510 (0.365 ~ 0.640)	-4.81 \pm 0.75	8.36 \pm 0.84	13.42 \pm 2.70	0.58 \pm 0.17	1.583 \pm 0.26
DM-C	12	22.57 \pm 3.23 *	0.348 (0.314 ~ 0.416)	-5.46 \pm 0.61 *	12.49 \pm 1.96 *	57.29 \pm 4.14#	0.89 \pm 0.29#	1.172 \pm 0.28#
DM-T	10	15.73 \pm 3.08▲	0.309 (0.333 ~ 0.383)	-4.97 \pm 0.29△	10.51 \pm 1.74△	47.31 \pm 6.09△	0.87 \pm 0.24	1.441 \pm 0.19▲

注：与 NC 组相比，* 代表 $P < 0.05$ ，#代表 $P < 0.01$ ；与 DM-C 组相比，△代表 $P < 0.05$ ，▲代表 $P < 0.01$ 。

表 17-8 17 周末各组大鼠脂肪组织 PPAR - γ 及 GluT - 4 蛋白水平的比较 ($\bar{x} \pm s$)

组别	鼠数	PPAR- γ (IA)	GluT-4 (ID)
NC	12	9986.85 (6599.68 ~ 14655.60)	1.640 (1.186 ~ 1.948)
DM-C	12	4750.70 (3346.63 ~ 10390.63) *	0.828 (0.621 ~ 1.010) *
DM-T	10	13613.35 (7293.85 ~ 17198.73) △	1.157 (0.921 ~ 1.953) △

注：与 NC 组相比，#代表 $P < 0.01$ ；与 DM-C 组相比，△代表 $P < 0.01$ 。

【分析与解答】 本例为内科学的动物试验，设计遵循了随机对照重复的原则，将受试对象随机分成了三个处理组。在方案中也给出了统计分析方法。但是只是笼统地说采用 t 检验或 z 检验或秩和检验，实际上 t 检验仅适用于两组呈正态分布且满足方差齐性的数据，在对各组大鼠相关生化指标的比较中，有 3 个组，组数大于 2，不适宜用 t 检验。因此，对于正态分布的指标应采用方差分析，方差分析有统计学意义时再做组间比较。而对于不服从正态分布的指标，就应采用秩和检验。检验有统计学意义的前提下再采用 Newmany 方法做组间比较。对于表 17-8 中两个指标的比较，存在同样的问题。

17.3 资料整理阶段常出现的问题

在进行一项科学研究时，非常重要的一点是，所有采用的统计分析方法都应该描述得十分清楚、详尽，以便专家或读者在接触到研究数据时能够验证所获得的结果。因此，在每一篇医学研究的论文中，凡是与统计相关的每一个细节，都应该对所采用的统计分析方法和技术交代清楚。

当使用多个统计检验方法时，重要的是要指出对于给定的数据集采用的方法是什么。如果检验方法存在配对或非配对情形，比如 t 检验，秩和检验等，应该指明所用的检验为哪种情形，究竟是配对还是非配对，同时还应说明所用的样本量是多少。

因为没有正确或清晰地表明所采用的所有检验方法，就会出现問題或错误。资料整理阶段可能出现的错误包括：

- (1) 没有指定样本含量是多少。
- (2) 没有说明检验是配对或非配对检验法，最常见的错误是不交代设计类型。
- (3) 统计检验的名称错误。
- (4) 对于并非常用的或者不清楚的方法没有给出具体的解释或者参考文献。
- (5) 用了多种检验方法时，没有指定对于给定的某一数据集所用的方法是什么。

17.4 数据呈现阶段常出现的问题

一项良好的科学研究，数据的呈现与数据的收集和分析一样重要。当采用统计方法描述或者表达所研究资料时，应该注意采用正确的统计方法描述集中趋势和离散程度。如果使用均值和标准差，那么显而易见该数据应至少是近似正态分布且无偏度的。否则这些描述的指标毫无意义。对于有偏度的数据，通常情况下，如果使用中位数、百分位数来描述集中趋势和离散程度，则使用中位数表达平均水平，用四分位数间距描述离散程度。

均数的标准误 (SEM)，虽然在统计描述中应用非常广泛和普遍，但它并不是统计描述的指标，它在统计描述中大量应用的原因似乎是因为它使得数据的变异看起来更小。但它并不是描述性指标，相反，却是用于统计估计的推断性指标。因此，用均数来描述正态分布资料的集中趋势是正确的，但是采用标准误指标描述离散程度却是错误的。也就是说，我们不适合用均数 \pm 标准误的形式来对正态资料进行统计描述。

对于主要终点变量或者主要的结果变量，只要有可能，就要给出相应的置信区间，因为单独的一个概率 P 值并不能给出效应值大小的实际意义，因此，统计方法和技巧的正确运用可以让读者增强对该研究的理解，获得更多关于该研究的信息量。如果是采用统计推断的方法在不同组间进行比较，则应该给出组间差别的置信区间而不是单个组的置信区间。报告时应准确给出具体的 P 值，而不是仅仅给出与检验水准值的比较，比如，“ $P < 0.05$ ”或“ $P > 0.05$ ”等。

以下总结了数据呈现阶段经常出现的错误和缺陷，以帮助医学研究人员在研究中避免：

- (1) 图表中的结果表达不恰当、不清晰或者错误。
 - (2) 有描述集中趋势的指标(比如均数)，但是没有描述离散程度的指标(比如标准差)。
 - (3) 对于给定的数据的离散程度，给出的是标准误 (SE) 而不是标准差 (SD)。
 - (4) 利用均数和标准差描述非正态分布资料。
 - (5) 描述变异时未指定 \pm 符号。
 - (6) 只给出了 P 值，未给出置信区间或者只给出 $P > 0.05$ 或者 $P < 0.05$ ，但是没有给出具体的 P 值是多少。
 - (7) 只给出某一组的置信区间而不是与对照组差量的置信区间。
- 由于本阶段的错误比较明显，在此就不再举例说明了。

17.5 数据解释中出现的错误

在科研的最后阶段，解释数据分析的结果是至关重要的，首先应该确保采用了最合适的和检验效能最高的统计分析方法，这是最基本的，这可避免研究的数据不支持研究结论的缺陷。如果声明

有统计学意义,那么一定保证已经采用了假设检验方法。如果结果没有统计学意义,那么在下结论的时候就要十分小心。因为缺乏统计学的差异并不意味着一定没有实际意义。例如,也许仅仅是样本量太小,没有足够的效能检验出统计学差异。也许结果中可能包含重要的临床发现或者对研究人员来说值得利用的结果。当采用了小样本,并获得了无统计学意义的结果时,对Ⅱ型错误(即假阴性错误)进行分析讨论是非常有必要的;同样,如果存在多个组间的两两比较,对假阳性错误的讨论有同样重要的意义。如果一个研究中潜在的混杂因素或偏倚未得到有效调整,解释结果时也应该实事求是地陈述。

(王素珍 曹文君 李炳海)

第 18 章 内科论文写作要领及注意事项

18.1 内科论文常见的写作问题

内科学医学论文的主要结构包括题目、摘要(关键词)、引言、资料与方法、结果、讨论和参考文献。

18.1.1 题目常见写作问题

题目应该简短明了、开门见山,对读者具有吸引力,能准确地概括论文内容,点明主旨,做到表里如一,题目一般不超过 20 个字。在题目的写作中应注意:尽量避免用虚词、开头避免使用阿拉伯数字、少用或不用副标题、尽可能避免使用标点符号。另外,需要特殊说明时在文题右上角标注“*”并在本页左下角说明,如*国家自然科学基金资助项目(编号 1234567);非通用的名词术语要书写全称不能随便缩写,如把甲状腺功能减退性心脏病写成甲减心等。

18.1.2 摘要常见写作问题

摘要是论文的缩影,是全文的概括和浓缩,是论文内容不加评论和注释的简短陈述,是一篇有数据、有结论的短文。它可以使读者一目了然,从而大大节约阅读时间,能在短时间内让读者决定是否值得进一步详读全文。目前,医学论文的摘要大多采用结构式摘要,其格式包括目的、方法、结果、结论四个要素。摘要的字数不宜超过 250 个实词,除特殊情况外,不用图表、化学结构式和非通用的符号和术语缩略语略称代号;在首次出现处必须加以说明;一般比较重要的论文都有摘要,内容较简单或较短的论文也可没有摘要。

在提到摘要时,不得不提一下关键词。关键词指用以表示论文主题内容的核心词语,它可以有以下作用:供读者了解全文涉及的主要内容;供读者检索已发表的有关文章;编制个人检索卡片;供计算机检索收录文章用。每篇论文选取 3~8 个关键词;可从题目或摘要中选取能代表论文主要内容的有关词或词组作为关键词,这些词应与正式出版的主题词表或词典提供的规范词保持一致;如主题词表暂缺可用本学科当前常用词。作为关键词,为了国际交流,应标注与中文关键词对应的英文关键词。可供查阅主题词的工具书有:汉语主题词表(中国科技情报研究所和北京图书馆主编)及医学主题词注释字顺表(中国医学科学院医学情报研究所,1984 年)以及全国自然科学名词审定委员会公布、科学出版社出版的医学名词(一)(二)(三)等。

18.1.3 表格常见写作问题

在进行内科学论文写作时,经常用到表格。表格是简明、规范的科技语言,是医学论文的重要组成部分,具有对比鲜明、表达力强、易得要领、便于计算和分析等优点,已成为现代科技文献不可缺少的表述手段。然而,表格的使用也存在较多问题,质量亟待提高。

1. 表格的选择

编制表格时,应首先判别它的必要性:

(1) 如表格栏目中的内容基本相同或为相同类型的表格, 应尽量删除或合并。

(2) 如表的内容简单, 仅少数几个统计数字, 可用一两句话表达清楚的, 应以简要文字叙述。

(3) 如同时使用插图和表格表述同一内容, 应考虑选择哪种表述形式更为合理。通常, 强调事物的形貌或参量变动的总体趋势时, 以插图为宜; 相反, 对比事项的隶属关系或对比量的准确程度时, 则以表格为宜。

2. 表格的种类

(1) 文字叙述表: 表格以文字叙述为主, 多见于临床研究论文中各种病症的比较。采用文字表时应注意归类总结, 力求专业术语少而精, 并根据需要加用表线。

(2) 统计表: 是医学论文最常用的一种, 它在表达、积累、分析、比较资料方面都有着极为重要的作用。

3. 制表原则

(1) 自明性强: 表格是完整的、可独立存在的形象化语言。表格的内容应简洁直观, 以数字表达为主, 避免夹杂过多的文字, 给人以强烈的对比效果。

(2) 主谓清楚: 作者在精心设计表格时, 应力求科学、准确、一目了然。一个好的表格应具有语言学上的逻辑性, 即主谓清楚、层次分明、标目合理。

4. 制表要求

(1) 结构完整: 表格的组成要素如下所示。表格结构方面存在的问题常为表题与内容不贴切、过简或转行时未注意语气的停顿; 主谓语标目不准确或缺少主谓标目; 有的排列混乱, 合计项或谓语纵标目分层时未加横线等。

(2) 内容突出: 表格的排列应简单明了、内容突出。根据对主语标志处理的不同, 可分为简单表、分组表、复合表。复合表应用最为广泛, 但其自身对照表较难理解和编制。来稿中常见的问题是重点不突出、自明性差、重复或资料不全, 令人费解。

(3) 逻辑排列: 统计表中的主语一般指被说明的事物, 多为文字, 如组别、类型等。有时, 也可以是体现主语的时序, 如年份或数字组段、时间、分值。谓语一般指用以说明主语的指标, 多为数字以及体现谓语的成分, 如例数、百分数、标准差或标准误等。通常, 主谓语标目合起来可以构成一句完整的话。来稿中常见的问题是主语或谓语缺少或主谓颠倒, 例数或动物数写在主谓语标目后等。

(4) 对比鲜明: 作为分析用的表格, 最主要的作用就是进行比较。为了更好地达到此目的, 应对对比事项(组别、指标)进行必要、准确的归类, 按比较的需要靠近排列。来稿中有的表格虽列出了很多数据, 但由于标目不合理或未给出标目, 令读者难以理解, 使表格失去了应有的、清晰的逻辑对比功能。

(5) 表达准确: 如比和率的运用, 最常见的问题是不加区分, 简单地以“%”代标目或将比误为率。而有的表虽列出了百分比或百分率的数值, 但表题、表头、表身内均未列出总例数, 给读者比较分析带来困难。

5. 表内数字格式

一律用阿拉伯数字, 上下个位对齐。数字中如有“±”或“~”号, 则以其为中心对齐。数字为零的例数或百分数应分别写为 0 和 0.0, 未取得数据者以“…”表示; 未做者则以“—”表示。表内有效数字应一致。表内数字常见的问题是: “…”与“—”用法混淆或一律以空白表示, 不知其含义; 还有的区间数缺乏连续性; 同一指标的有效位数不一致等。

6. 表内单位格式

表格内数据的单位有公用单位和特有单位。公用单位可直接写在表题后并加圆括号，特有单位可写在相应标目后并加圆括号，且表体内单位应与正文一致。单位中常见的问题是书写错误；计量单位列在表内数据后或不标示；体液检测值单位未换算成 L 而用 ml 或 dl 表示；时间单位用法不一，有的用中文，有的用符号。按中华医学会规定，表格中时间单位统一用符号表示，如伤后时间 (d)，而不写成伤后时间 (天)。作为分母时，原则上应与分子的计量单位表示法相同，即分子、分母的计量单位同用中文或同用符号，如次/日、r/min。

7. 表格的注释

表题、标目或某个数据需注释时，可在其右上角加注释符号，并在表下用相同的符号加注相应的文字。对表需做附加说明者，可在表下加“注：……”，句末可不用标点。个别作者投稿时由于未仔细核对，常常是表中数据标有*，表下却没有注释；或表下有注释，表内数据却无注释符号。

8. 表格中的统计符号

论文中的显著性检验，只在表下注释 *P* 值是不够的。应将检验方法、计算结果及 *P* 值均列出，以便读者进一步了解实际差异的大小。一般将统计结果的表示方法如 $\bar{x} \pm s$ 写在表题后面或谓语标目处。*t* 值或 *P* 值等作为标目词时不能只写“*P*”、“*t*”，而应加“值”。表下注释 *P* 值结果时，应注明比较对象 (如 $\times \times$ 与 $\times \times$ 比较, $t = \times \times$, $*P = 0.0258$)。我们发现不少来稿中表内数据为 $\times \times \pm \times \times$ ，却未注明是 $\bar{x} \pm s$ 还是其他内容；*P*、*t* 作标目时未加“值”字；表下注释 *P* 值结果未注明比较对象。动物数或例数均以 *n* 替代，按中华医学会规定应一律用中文 (如鼠数、犬数、例数) 表示。

9. 表格的编排位置

表格的位置应紧随“见表 \times ”或“(表 \times)”的文字的自然段落之下，即先见文，后见表。如作者将所有表格另纸放在最后，正文中也应以“表 \times ”标示其所在位置。

18.2 内科论文写作要求及注意事项

内科学论文的写作即将有目的观察病例所获得的数据结果或临床积累的资料通过科学的思维判断推理用文字图表符号等再现的过程。不管你从事任何临床工作，只要你希望把自己的研究成果或临床经验以文字的形式进行交流和发表，就不能不涉及医学论文的写作。

18.2.1 写作要求

1. 医学论文写作的基本原则

一篇好的临床医学论文必须具有科学性、创新性、实用性、条理性和规范化，这是撰写医学论文必须遵循的基本原则。

(1) 科学性。科学性是指论文资料翔实、内容新颖，必须是客观存在的事实且能经得起科学的验证和实践的考验，并具有可重复性。应做到立论客观、论据充足、论证严谨。对临床观察和实验所得的数据和实践材料加以整理分析和推导不能主观臆断，更不能为达到预期目的而歪曲事实伪造数据。具体要求是：数据资料来源正确可靠，如对病例来源及其选择应有一定标准；要有周密的临床科研设计；对观察病例要进行随机分组；试验组与对照组要具有可比性；观察指标的选定必须客观；统计学方法应运用恰当。

(2) 创新性。创新性是指论文总结的成果有新的发现、新的理论、提出新的设想、探索新的规

律;采用的是新方法或新技术。在重复模仿他人已有成果或科研课题中补充实现该成果的新条件、新方法或新改进即仿中有创。

(3)实用性。实用性是指通过基础或临床医学的科研活动解决医学实践中存在的实际问题。一般情况下,临床医学论文对临床医学实践具有直接的指导意义,其实用性较强。不论其报道的是成功的经验还是失败的教训,都可为他人所利用或借鉴。有些研究虽然暂时还不能解决实际问题,但从发展眼光来看仍有其重要意义者也应列入有实用价值的范畴,如回顾性研究可为前瞻性研究提供一些有价值的参考资料。

(4)条理性。条理性是指思想语言文字达到的层次要求:用客观的论据和符合逻辑的推理来论证和阐述问题形成结构严密、论点鲜明、论据充分、论证有力、结论明确、说理性强、可信度大的文章;文字简洁、语法正确、表达清晰、层次分明,读起来流畅易懂,避免使用华丽的辞藻和夸张性词语。

(5)规范化。规范化也是医学论文写作的基本原则之一,一篇好的医学论文要按规定的格式撰写,除了在内容发表上可以达到要求外,还要保证论文的写作质量,还应该在名词术语、计量单位、数字用法、外文字母和规范汉字上达到规范化的要求。名词术语应该使用全国自然委员会(1987年由国务院批准成立)审定的名词。1987年至今已正式公布的科技名词出版了近30种,其中与医学有关的名词约有20种。

2. 资料与方法

资料中应包括下列内容。病例来源:观察对象来自某年某月,是住院还是门诊患者,分组是多少例。完成疗程的有多少例。一般资料:包括年龄、性别、病程、就诊时间、伴发疾病等。诊断标准:为了方便别人参考诊断标准,若属于国际或国内标准的则给出参考文献。排除标准:指出哪些情况不在本次研究之列。分组情况:要随机分组,不但要介绍治疗组的情况,还要介绍对照组的年龄、性别等,更要说清楚两组是否具有可比性。统计学方法:对不同的数据资料(计数资料、计量资料、有序资料、名义资料)要选择正确的统计学方法。治疗方法应具体陈述:如观察药物治疗某种疾病时,应介绍药物剂量、剂型、服用方法、疗程等。实验室检查及特殊检查:一些常规检查包括尿尿便常规检查、X线胸透、心电图等;特殊检查包括脑电图、CT、核磁共振等;随访情况包括随访人数及失访人数、随访结果、有无误诊等;实验器材要写清名称、规格、性能,以便别人能够知晓或检查或选用。

3. 结果

结果是论文的核心部分,是作者将研究过程观察所得的原始资料或数据经过审查、核对、分析、归纳和正确的统计学处理后得出来的结论。未经统计学处理的观察记录数据叫原始数据,是不准确的数据,由此得出的结论是不可靠的。统计学处理主要是使原始数据从难以理解变得容易理解,并从原始数据的偶然性中揭示出隐藏在其中的某种必然规律。结果若能用图表的形式表达清楚就不必赘述文字,图与表之间的内容不能重复。

4. 讨论

讨论是论文所要报道的中心内容,是将研究结果从感性认识提高到理性认识的阶段,它应该包括:对所得结果进行补充说明或解释;重点说明该项研究的创新性、先进性;对结果进行分析探讨,并对可能原因机制提出见解或阐明观点;将结果与当前国内外研究结论进行比较,并提出新的见解及对其理论和实践意义做出评价;提出在调查研究过程中的经验体会;指出该结果的可能误差。此部分的写作要求是:大量阅读有关文献,充分地了解本研究的历史发展及现状;不罗列文献;详略得当、重点突出、着重关注围绕本研究课题近5年文献的结果和观点;比较自

己的结果与别人的异同点；实事求是地对自己的研究成果进行客观评价；对于与别人不同的研究结果要分析原因，切忌报喜不报忧；对某些现象不好下结论时，措辞要客观，留有余地，不要轻易去填补；层次应清楚，如讨论的问题较多可按内容进行分段列出小标题，每段围绕一个论点加以论证，做到论证充分。

5. 参考文献

参考文献是指为撰写或编辑论文引用的有关图书和期刊资料。引用参考文献时，应按其在文内出现的顺序依次以阿拉伯数字在文后标注。参考文献引用得是否正确，书写格式是否符合标准已成为审核医学论文和医学期刊质量的一项重要指标。选用参考文献的时候需要经过筛选。列出参考文献的目的是：为了说明本文所借鉴内容的科学依据，列出出处以供读者查阅参考；减少对前人文献的复习以缩减篇幅；认同与尊重前人成果及著作。引用参考文献应遵循以下几个原则：权威性 & 专业性，权威性专业性杂志上发表的科技论文最能代表相关课题的研究水平及现状，应尽量引用；自阅性，即所引文献必须是作者亲自阅读过的文献；准确性，即正文中标注的准确性和著录项目的准确性；公开性，即公开发表的论文不应引用非公开发表的文献。

18.2.2 注意事项

在内科学医学论文的结构中，摘要部分一般有固定的格式，即“目的、方法、结果、结论”，摘要部分只求简单扼要，不要详细的解析和分析。因此，下面重点讨论 4 个应注意的问题。

1. 引言

引言俗称为“开场白”，对判断文章的价值有重要的作用。它既给读者，也给审稿人和期刊编辑一个“先入为主”的印象。临床文献评判提纲的第一条就问道：“作者研究的目的是什么？是否为解答某个（些）研究问题或验证某个假说？如果是的话，你能否从文章中归纳出该研究的问题或假说？这个研究问题或假说是否关于某个干预（如治疗）的效果、因果关系或确定某个健康问题的强度？”引言中最重要的内容就是要清楚地将研究的目的告诉读者，同时要交代研究的背景和立题的依据。一般要求先从总体上介绍研究背景，逐渐缩小到立题依据，再集中到研究目的。

国际上有一些医学期刊，要求在引言部分详细地介绍研究背景和立题依据，但更多的国际和国内医学期刊，是要求在引言部分简明扼要地介绍研究背景和立题依据，然后集中到研究目的。这样的话，一些研究背景和立题依据的内容，就需要在讨论中描述。一个好的引言应该是先用几句话概括出先前这个领域的研究，然后再说明这个研究的必要性，必须如实地阐述你的研究与先前的研究相比，创新性在哪里。

2. 材料与方法

(1) 材料与方法这一部分的内容往往体现出文章的科学性和有效性，因此它是期刊确定论文取舍最关键的所在。

目前，内科学的论文普遍存在这样的问题：如果论文涉及实验内容，则将大部分的篇幅放在描述实验室的方法上，而忽略了体现文章的科学性和有效性的内容；如果论文不涉及实验内容，则不知道这部分该写什么，不少临床论文在“材料与方法”部分中，只用几行字，因为“没东西好写”。其实，一篇论文的质量高低，往往体现在这一部分中。

(2) 不少论文将病例的性别、平均年龄、平均病程、基线的比较等“结果”部分的内容提前写在“材料与方法”中，这是不恰当的。“材料与方法”主要是描述该研究的计划，主要包括以下 6 个方面。

① 研究策略：论文需要阐明研究策略，可以直接，也可以间接，但必须让读者很明确地知道本研究的策略。

② 研究样本：临床研究的样本主要是病人，样本来源必须被阐明。研究病例最好是某一时段内的连续性的病人，如果不是连续性的病人，有必要阐明如何从大组的病人中选取研究样本。研究样本的入选标准和排除标准也必须详细描述。在病例对照研究中，对照组的构成和来源必须交代清楚。

③ 研究因素：在实验性研究和分析性研究中，研究因素非常重要，往往是该项研究的核心内容。因此一定要明确交代研究因素以及测量这些研究因素的方法。在实验性研究，要明确分组是否真正随机，如果不是随机，必须比较两组间的基线是否基本相同；干预组和非干预组的治疗方法有何区别，除干预之外，其他治疗是否相同，干预组和非干预组各自的治疗方法必须交代清楚。在分析性研究中，研究因素存在与否，是病人进入研究时已经存在，而不是按照研究者的意愿进行分组的。因此，需要交代的是其测量方法，而不是分组。在队列研究，研究因素的确定在先，结局的判断在后，一般较少发生研究因素测量时的主观偏倚；但在病例对照研究中，由于先确定病例组和对照组，在回顾暴露因素时容易出现主观偏倚，论文在这里必须交代避免偏倚的方法。

④ 研究结局：研究结局的判断必须准确无误，因此需要交代判断研究结局的方法。如果结局的判断受研究者主观因素的影响，如关节炎的放射学损害，则必须交代是否采用盲法。在随访研究（包括随机临床试验和队列研究）中，失访病例如何归属，及其随访时间如何确定，均需交代清楚。对于不以死亡为结局的研究，如何处理死亡者的资料，是删失还是终点，时间如何确定等需交代明白。

⑤ 控制混杂和偏倚的措施：如果研究因素和研究结局之间有可能存在混杂因素，则必须有措施去校正混杂因素所导致的偏倚。例如，对于多因素的因果关系推导，必须采用多元回归的方法校正各因素之间的相互混杂。减少或避免偏倚的措施需要交代，如在时间较长的随访研究中，如何减少失访。盲法不单只应用于实验性研究，在分析性研究中，研究因素和研究结局的确定和判断，也常常需要采用盲法。

⑥ 统计学方法：实验性和分析性的医学论文，均必须明确地交代统计学方法。有不少论文没有交代统计学方法，如某些含有实验研究内容的文章，由于花了大量的篇幅介绍实验流程，使“材料与方法”部分显得“充实”，因此遗漏或省略了统计学方法的描述，使文章不完整和欠缺科学性。有些文章在“统计学方法”一栏中，只用了 5 个字：“采用 t 检验”；也有些只写“全部数据均由 SPSS 13.0 软件进行统计学处理”，没有交代具体的统计学方法。让人感觉到，这是在应付式的填补“缺项”，写了等于没写。虽然 t 检验和 χ^2 检验是基本的统计学方法，但并非“放之四海而皆准”。首先，并非所有资料的统计分析均可以用 t 检验和 χ^2 检验，病例数少和非正态分布的资料就必须用非参数检验方法，如秩和检验和精确概率法等。多组间的比较可能需要采用方差分析；随访资料（如随机临床试验、队列研究等）用生存分析方法可能更佳；存在多因素的因果关系分析（如预后因素分析、病因学研究等）需要多元回归分析，等等。

另外，样本量的推算主要是使研究的结果避免犯统计学第Ⅱ类错误。第Ⅱ类错误是指由于样本量不足，使得本来有差异的两组，运算出无差异的结果。样本量的推算是根据研究者期望的差异值、检出意义的水平（ α 值）和研究的功效（ $1 - \beta$ 值）而来，其中 β 值是允许犯第Ⅱ类错误的概率。一般要求 $\alpha = 0.05$ ，研究功效是 0.8 或 0.9。当你的研究结果中存在 $P > 0.05$ 时，则要求文章有样本量的推算，或提供研究功效。

3. 结果

写论文最主要的目的就是向读者展示研究的结果，论文的核心和精髓主要体现在“结果”部分。因为你在“引言”部分所提出的问题，均可在“结果”部分找到答案。这部分要求对研究结果进行客

观的陈述,既不受其他文献的影响,也不受作者主观意愿的左右。所以“结果”部分不需要拓展,只是要客观的陈述。

一份好的“结果”,必须让读者顺着你的思维,一步一步地获得答案。通过文字告诉读者本研究发现了什么;通过表格将具体的数据有条理地展现给读者;通过图形使读者能够直观地理解研究的结果。

(1)虽然有些作者将研究资料基线的描述放在“材料与方法”中,但实际上,基线的描述放在“结果”中更加确切。因为“材料与方法”主要是交代研究的计划,至于所收集的病例中年龄多大、男性占多少、病程多长等,是在结束研究后资料分析时才知道的。如果是分组的研究,需要比较各组间的基线是否相同。随访研究还要比较失访者或中途淘汰病例的基线是否有别于组内待分析病例,让读者知道失访或中途淘汰病例是否导致研究的偏倚。

(2)实验性和分析性研究的结果很大部分是来自统计学运算的结果。

① P 值问题。尽管我们一直强调“应尽可能给出具体的 P 值”,但论文中用“ $<$ ”或“ $>$ ”来表示者仍占绝大多数。既然文章中已阐明统计学运算是采用统计软件,计算机运算的结果就肯定是具体的 P 值,因此论文没有理由不给具体的 P 值。只有当 P 值非常小,计算机运算结果显示的 P 值为0.000”时,论文中才标“ $P < 0.001$ ”。

②注意相关强度及其95%可信区间的表示。临床研究的结果常常涉及相关强度的运算,最常用的是OR值和RR值。有些已发表的论文,作者明明在统计学方法中说用logistic回归或COX回归,但结果中却找不到OR值或RR值。另外,值得一提的是,不少文章中OR值或RR值的95%可信区间包含了1, P 值却 < 0.05 ,计算机肯定不会运算出这样的结果。

③数据必须与研究资料一致。在少数论文的结果中,数据与研究资料相差太大。例如,一篇长期随访的研究,在“材料与方法”中注明为“1985年1月至2000年12月”,而“结果”中的3个生存曲线图,均显示曲线右端到达25年,究竟是哪个环节出了问题?

④避免违反科学的数据。不时见到一些违反科学的数据,使论文成为“垃圾”。例如,一篇狼疮性肾炎长期随访的论文,将1组病人的两个结局(人生存率和肾脏生存率)的2条生存曲线描绘在一个坐标上,运用log-rank检验,竟能“计算”出其 χ^2 值和 P 值。计算机是不可能做这种运算的,因为这两条曲线不可能做差异性检验。

(3)凡是可以进行统计学处理的资料,都应尽量运用统计学处理。但有些资料则需要进行详细的陈述。应该说,该研究资料所能反映的各种结局都应该被陈述。例如,淘汰或失访病例的数量和主要原因;各种疗效和各种副作用;研究中出现的各种问题,包括意外的发现等。有些意外事件的发生虽然不一定与本研究有直接联系,但也必须如实描述,然后在“讨论”中进行分析,陈述自己对这意外事件的观点。陈述必须客观,不要牵强,也不要讨论和主观推断,讨论和推断必须放到“讨论”中去。

4. 讨论

讨论部分呈梯形结构,提示这一部分需要根据研究的结果,结合文献的回顾,拓展开来分析讨论,以得出本研究的结论。

如果引言部分限于篇幅,无法详细介绍本项研究的背景和立论依据,则可以在讨论的第一段补充描述。需要注意的是,论文中不要輕易出现“首创”、“领先”或“过去未见报道”的字眼。如果确是如此,应该客观地陈述:用哪些词(如甲氨蝶呤和成人Still病)?检索哪个数据库(如Medline或PubMed)?检索的结果如何?等等。因为从这些数据库中找不到的文献,不等于全世界就没有人报道过。客观地陈述检索结果,容易被人家接受,而且读者必要时也可以运用同样的策略,重复进行检索。

讨论中必须紧紧围绕着自己的研究结果,分析国际上其他学者同类的研究,哪些是相同的,哪些是不同的甚至相反的,并分析不同的和相反的原因。如果你的资料是真实的,研究设计是科学的和合理的,就应该敢于相信,并坚持自己的观点。科学的东西,总是会存在不同的观点。当然,文章中有必要分析其原因,如种族的差异、研究样本中亚型的不同等。讨论中需要避免离开自己的研究结果,讨论得太远,导致离题。临床文献评判提纲的最后一条就问道:“研究结论的引述是否根据研究分析的结果?针对研究的问道,作者做出了什么结论?根据本研究,作者是否又提出了新的假说?”一篇临床医学论文,往往就是围绕着这些问题画上句号的。

(安洪庆 李炳海)

参 考 文 献

- [1] 陆再英,钟南山. 内科学(第7版)[M]. 北京:人民卫生出版社,2008:121-123.
- [2] 王汝芬,陈景武,李建国等. 医药科学研究统计方法[M]. 山东:山东科学技术出版社,1994:1-17.
- [3] 杨锡强. 我国儿科临床医学研究的现状与思考. 中华儿科杂志[J], 2008, 46(5):321-4.
- [4] 胡良平主编. 检验医学科研设计与统计分析[M]:北京:人民军医出版社,2004:321-4.
- [5] 胡良平主编. 口腔医学科研设计与统计分析[M]:北京:人民军医出版社,2007. 321-4.
- [6] 陈峰. 医用多元统计方法[M]. 2版. 北京:中国统计出版社,2007.
- [7] 徐勇勇主编. 医学统计学(第二版). 北京:高等教育出版社,2004.
- [8] 孙振球主编. 医学统计学(第三版). 北京:人民卫生出版社,2010.
- [9] 方积乾主编. 卫生统计学(第六版). 北京:人民卫生出版社,2007.
- [10] 李立明. 流行病学(第六版). 北京:人民卫生出版社,2007.
- [11] 颜虹. 医学统计学. 北京:人民卫生出版社,2005:329-338.
- [12] 金丕焕. 医用统计方法(第二版). 上海:复旦大学出版社,2003,126-132.
- [13] 易东. 军事医学统计学. 北京:人民军医出版社,2004:194-206.
- [14] 楼东华. 引入协变量比较药物的效果. Journal of Mathematical Medicine, 2000, 13(1):66-67.
- [15] 马斌荣. 医学科研中的统计方法. 北京:科学出版社,2001,87-99.
- [16] 李河. 协方差分析的正确应用. 循证医学, 2004, 12, 4(4):224-227.
- [17] SAS 9.2 Macro Language Reference, SAS Institute Inc. 2008.

第 19 章 临床科研统计设计要览

在临床科研设计中，必须牢牢抓住以下几个核心内容：熟悉相关的政策与法规、严格遵守伦理道德、降低心理因素对结果的影响、尽可能提高受试者的依从性、全面把握三要素、认真落实四原则、合理选用设计与比较类型、实时并严格控制临床试验研究质量、正确无误地在临床病例报告表上进行记录。

19.1 临床科研的种类

19.1.1 调查某些疾病的发病率或人群分布规律

临床工作者常需了解在一定范围和时间内某种疾病的患病率、发病率、治愈率和死亡率，以便为国家或省市级卫生防疫部门提供一些准确数据，这不仅有利于了解一定区域内的人群在某种特定的疾病方面的患病或发病或治愈或死亡情况，更有利于制订出科学完善的卫生政策、法规和卫生资源的合理配置方案，指导卫生防疫、疾病筛查、预防、诊断和治疗等工作持续有序地进行，确保人民身心健康，提高人民身体素质。事实上，为了提高调查研究的效率，每个调查研究项目一般不会仅调查一种疾病的患病率与发病率，可能会同时调查多种相关疾病的患病率与发病率，如同时调查某省某年 5 种或 10 种特定的恶性肿瘤的发病率、患病率、死亡率、病死率、生存率。

除了调查某些疾病的患病率与发病率等统计指标外，研究人员还常会调查某些疾病的发病原因、人群分布规律与特点。因为前者仅仅是对人群健康状况的一个非常粗略的描述，对疾病的发生频率有一个“点估计”；但对导致疾病发生的原因、对疾病的人群分布规律与特点、对诊断与治疗疾病、对预防疾病等都关系不大。因此，调查研究的内容需要进一步扩充和丰富。一旦找准了疾病发生的原因，有利于采取积极措施消除其影响，可以有效地预防疾病，达到明显降低患病率与发病率之目的；一旦掌握了疾病的人群分布规律与特点，也有助于找准疾病的发生原因或提高对疾病的诊断水平或对症治疗的策略。

上述各项研究常被称为流行病学调查研究。相对于在医院所做的诊断与治疗疾病的研究而言，称上述研究为开放性“临床研究”较为合适。

19.1.2 为某些医学指标制定参考值范围或为某些疾病的诊断与治疗制定标准

无论是正常体检还是患者到医院去就诊，医生常会给受试者出示很多检查结果，如血小板、红细胞、白细胞、尿酸、血沉、谷丙转氨酶、谷草转氨酶、乙肝表面抗原……在这些指标的测定结果右边会列出两个数，一个较小 (r_L)，一个较大 (r_U)，医学上称这两个数据为该定量指标的参考值范围，通常为 95% 正常值范围。其含义是：在总体中，约有 95% 的个体的定量指标的测定值落于区间 $[r_L, r_U]$ 内。那么，此 95% 参考值范围是怎样获得的呢？获得的方法如下：先确定一个具有同质性（人群的年龄范围、身体健康状况等必须符合事先规定的标准）的总体，再从该总体中按一定规则（等概率地还是按某些重要非试验因素分层）进行随机抽样，样本量应足够地大（应根据特定公

式计算而得 N 的数值), 用准确的方法测定样本中每一位个体该定量指标的取值, 然后, 根据数据的分布规律(正态分布或非正态分布), 采用合适的统计分析方法(常分为正态分布法或百分位数法)计算出一个下限值和一个上限值, 它们就是所求的 95% 参考值范围。完成此项工作的全过程称为确定某医学指标的参考值范围的研究。

很多人即使患了同一种疾病, 有时, 他们的症状或体征可能相差很多, 此现象被称为“个体差异”。生物体之间的一个明显特征就是存在极大的个体差异, 这增加了人类对很多问题的研究的复杂性。为此, 有必要让很多有多年临床工作经验的医生参与讨论、甚至通过精心设计的试验, 对某些疾病的特征和治疗方案进行全面而系统的论证和总结, 以便制定出对此类疾病准确诊断和有效治疗的标准, 为技术水平薄弱的医院和医生诊断与治疗此类疾病提供技术支持和保障, 将有利于提高人类诊断与治疗此类疾病的水平。此类研究属于为某些疾病的诊断与治疗制定标准。

19.1.3 为评价某治疗方法或药物或医疗器械的有效性与安全性

为评价某治疗方法或新药或新医疗器械或某老药治疗新适应症的有效性与安全性是临床科研最核心的内容。治疗方法通常分为一般药物治疗(包括中药、西药或中西药结合)、化学药物治疗(简称化疗)、放射治疗、手术治疗、器官移植等; 新药的临床试验通常包括 I、II、III 期, 对少量新药还进行 IV 期临床试验; 对新医疗器械一般也会进行 I、II、III 期临床试验; 有时, 人们发现使用了多年的某种药物可用于治疗另一种疾病, 为了证明这一发现具有普遍意义, 也需要进行临床试验, 属于老药治疗新适应症的有效性与安全性的评价问题。例如, 很多医生发现: “痢特灵”原本是用来治疗痢疾的药物。然而, 在用痢特灵治疗痢疾的同时却治好了溃疡病。为了弄清下面两个问题所进行的临床研究, 就属于“老药治疗新适应症的有效性与安全性的评价问题”。第一个问题: 痢特灵是否真能治好溃疡病? 第二个问题: 若能治好溃疡病, 是否能治好身体上任何部位上的溃疡病?

19.1.4 同类临床研究的 meta 分析

同类临床研究的 meta 分析, 是运用统计技术对多个同类的临床研究结果做概括的综合性分析, 以对药物的效应做出更稳健的估计、评估各个研究结果之间的异质性和一致性、分析不同类型的患者对所用疗法的效应、增强研究报告结论的说服力、确认进一步进行重点试验的必要性并进行良好的设计。

19.1.5 为探查某些疾病的发病原因的研究

很多疾病的治疗效果并不理想, 有些疾病还根本无法治疗。原因有两个方面: 其一, 有些疾病的发病原因很不清楚, 如很多恶性肿瘤; 其二, 有些疾病的发病原因已经清楚, 但尚未找到对症治疗的药物或疗法, 如艾滋病。第一方面的临床科研属于为探查某些疾病的发病原因的研究; 第二方面的临床科研属于寻找特效药或疗法的研究。

19.1.6 基因治疗研究

很多疾病的发生、发展都与基因有关, 当这些与疾病密切相关的基因发生突变时, 就导致相应疾病的发生、发展。临床研究的任务之一是要尽可能多地找出与各种疾病密切相关的基因, 并寻找合适的药物或治疗方法, 从基因水平上对疾病进行治疗, 此类研究被称为基因治疗。

19.2 临床科研的核心内容

19.2.1 临床科研需要关注有关的政策和法规

因为临床科研的受试对象主要是人(健康人和某病患者),这就不可避免地要涉及拟开展的临床研究项目是否“合法”和“国际准则”。因此,不同类型的临床科研项目,将会受到不同级别的政策和法规的制约。

为了确保受试者的权益与隐私受到保护,不受侵害,确保试验结果可信,数据可靠,结论科学有据,对新药、新医疗器械、新疗法或手术方法的安全有效性做出科学评价,需要制定相应的法律法规。表 19-1 给出了中国、美国、欧盟和日本的管理当局、具体管理部门所制定的主要相关法规。

表 19-1 中国、美国、欧盟和日本的管理当局、具体管理部门所制定的主要相关法规

地区	管理当局	管理部门	主要法规
中国	国家食品药品监督管理局	国家食品药品监督管理局注册司	《中华人民共和国药品管理法》 1984、药品临床试验管理规范(GCP)1999、《药品注册管理办法》 2005
美国	Food and Drug Administration(FDA)	Center for Drug Evaluation and Research(CDER)	Food, Drug and Cosmetic Act, 1938; Code of Federal Regulations(CFR) title 21
欧盟	The European Agency for the Evaluation of Medicinal Products(EMA)	The Committee for Proprietary Medicinal Products(CPMP)	Directive 65/65/European Economic Community (EEC), 1965
日本	Ministry of Health and Welfare (MHW)	Central Pharmaceutical Affairs Council (CPAC) and the Pharmaceutical and Medical Device Evaluation Centre (PMDEC)	The Pharmaceutical Affairs Law, 1960

19.2.2 临床科研中不可回避的伦理道德问题

在以人为受试对象的临床研究中,必须妥善解决好伦理道德问题。具体体现在两个方面:其一,拟开展的临床研究项目是否得到了有资质的伦理委员会的审批;第二,临床研究项目的负责单位和负责人是否与受试者签订了知情同意书。

临床试验应遵守有关的法规和指南,如药品管理法、药品注册管理办法、新药审批办法、药品临床试验管理规范(GCP)等。同时,所有以人为受试对象的研究必须符合《赫尔辛基宣言》和国际医学科学组织委员会颁布的《人体生物医学研究国际道德指南》的道德原则。2008 年 10 月,第 59 届世界医学大会通过《赫尔辛基宣言》修正案。修正案扩展了宣言的适用对象,重申并进一步澄清了基本原则和内容,加强了对受试者的权利保护,同时还增加了临床试验数据注册和使用人体组织时的同意等新内容,提高了人体医学研究的伦理标准。

进行药物临床试验必须有充分的科学依据。进行人体试验前,必须周密考虑试验的目的、要解决的问题、预期的治疗效果及可能产生的危害。预期的收益应超过可能出现的损害。选择临床试验方法必须符合科学和伦理标准。

临床试验开始前,试验方案须经临床研究负责单位的国家临床试验机构伦理委员会审议同意,

并签署批准意见后方可实施。在试验进行期间,试验方案的任何修改均应经伦理委员会批准后方可执行。试验中发生任何严重不良事件,均应向伦理委员会报告。应急信件也正是从医学伦理角度充分考虑受试对象的利益而设置的。

在药物和医疗器械临床试验的过程中,必须对受试对象的个人权益给予充分的保障,并确保试验的科学性和可靠性。研究者必须向受试对象提供口头或书面的有关临床试验的详细材料,包括试验目的、预期收益、受试对象被分配到不同处理组而可能发生的风险与不便、因参加试验而受到损害或影响身体健康时能够获得的治疗和补偿。研究者不能强迫受试对象参加试验,经受试对象同意后,须有受试者或其法定代理人在知情同意书上签字并注明日期,执行知情同意过程的研究者或其代表也需在知情同意书上签名,并注明日期。知情同意书应使用受试对象能够懂得的语言和文字,受试对象保留在任何时候退出试验的权利。

19.2.3 临床科研中应尽可能降低心理因素对结果的影响

有研究表明,安慰剂的效应可以高达 50%,甚至 70% 以上。这说明:在临床科研中,来自研究者和受试者的心理因素的影响是不可小视的。而消除或降低心理因素对临床研究结果影响的最有效途径是采取合理的“盲法设计”。

盲法(blind method)作为纠正偏倚的一个重要措施,特别在临床试验中,是为了避免研究者与受试者的主观因素(特别是心理作用)对试验结果的干扰。临床试验根据设盲的程度分为非盲(开放性试验)、单盲、双盲和三盲。除开放性试验(open label)和某些不宜设盲的试验如外科手术、引起生活方式改变等干预试验外,一般均应采用盲法。仅有受试者不知自己所接受的是何种疗法时,称为单盲(single blindness);而参与临床试验和判断疗效的医护人员与受试者都不知受试者所接受的为何种疗法时称为双盲(double blindness),从而使研究结果得以客观评价;三盲(triple blindness)则是研究者(包括医护人员)、研究对象和负责资料收集和分析的人员均不知受试者所接受的为何种疗法,从而更好地避免了偏倚。

19.2.4 临床科研中应尽可能提高受试者的依从性

临床研究项目通常都需要一定时间长度才能完成,在此期间,可能会出现某些客观原因(如下雨下雪、意外事故、局部地震等)使部分受试者不能按时到指定地点接受访视(治疗或测量);也可能某些受试者对治疗效果的期望值过高或过快而对当前的临床试验不满意;还有可能因为研究者(通常为医护人员)的态度或心情等原因导致部分受试者不愿配合临床试验。总之,可能存在多种诱因导致受试者违反临床试验设计方案的要求或提前离开临床试验研究项目。这种结果可被统称为受试者依从性差。在一个临床研究项目中,若依从性差的受试者所占的比例过高(一般不应超过 20%),将直接导致此项临床研究项目的失败。提高依从性的措施如下。

(1)坚持患者自愿的原则,患者进入临床试验,应签订知情同意书,由试验者和患者共同签字。在签订知情同意书之前,试验者应向患者交代试验目的、要求、所获个人的利益及可能出现的毒副反应等。

(2)在患者进入试验时,对患者详细说明治疗方案为每一个患者提供一份说明书。

(3)在计划研究方案时进行预试验,对患者的依从性进行评价,并根据预试验结果,对方案做必要调整,以提高患者的依从性。

(4)治疗方案采用患者易于接受的剂量、剂型等,有利于提高患者的依从性。

(5)对长期随访患者要确定一个适当的随访间隔,既不能太长,也不能太短。

(6)必要时,对患者的尿、血中的药物浓度进行测定检查。

(7)做好一切可能的应急预案,尽可能降低客观原因对依从性造成的不利影响。

19.2.5 临床研究中应全面把握试验设计的三要素

试验设计的三要素(受试对象、影响因素和试验效应)在临床试验研究中非常重要,必须全面把握、认真落实。与受试对象有关的要点是:受试对象种类的确定(即什么疾病的患者)、受试对象的质量标准的制定(通常应制定出五个标准,即纳入标准、排除标准、中止标准、终止标准、剔除标准)和受试对象数量的确定(有根据地估算出合适的样本含量)。与影响因素有关的要点是:依据研究目的和条件确定试验因素及其水平,找准找全对主要评价指标有影响的重要非试验因素并设法对其进行有效的控制。与试验效应有关的要点是:选准诊断、疗效评价和安全性评价三方面的主要指标与次要指标,确定指标的测定方法、测定时间和测定次数。三要素的质量将直接影响临床研究结果的可信度。

19.2.6 临床研究中应认真落实四原则

试验设计的四原则(随机、对照、重复和均衡)在临床试验研究中非常重要,必须全面把握、认真落实。与随机原则有关的要点是:千万不要以“随便”取代“随机”。在样本含量(小于1000例)并不很大时,有时,完全随机的效果并不一定很满意。比较好且又具有可操作性的做法是,针对个别重要非试验因素采用分层随机化。与对照原则有关的要点是:不设对照组的临床研究是没有说服力的。出现对照不全或对照不合理同样也是不好的。设置合理的对照组是临床研究的一个重要内容,其合理性应体现在基本常识、专业知识和统计学三个方面。与重复原则有关的要点是:不应根据研究单位自身的条件随意确定样本含量,必须结合具体的临床研究项目的规模、性质和标准,有根据地估算出合适的样本含量,应特别强调每个小组中的样本含量,而不是单看整个临床研究项目中的全部样本含量。与均衡原则有关的要点是:均衡原则本身并没有任何可操作的内容,它仅仅是一个“验收”或“把关”的原则。一个临床研究项目的均衡性,与临床研究实施过程中的每一个环节都有关系,每个环节都做得尽善尽美,就是均衡原则的终极目标。

19.2.7 临床研究中应合理选用设计与比较类型

试验设计类型决定了一个研究项目中因素及其水平的一种结构关系和内在联系,而比较类型是与分析目的有对应关系的统计分析方法,二者不可混淆。设计类型通常分为单因素设计与多因素设计两大类,而单因素设计又常分为单组设计、配对设计、成组设计或称为单因素两水平设计和单因素多水平设计;多因素设计本身并不是一种具体的设计名称,它常分为两大类,即非回归设计与回归设计。

在非回归多因素设计中,常分为无法考察交互作用的多因素设计(如无重复试验的随机区组设计、无重复试验的平衡不完全随机区组设计、无重复试验的拉丁方设计、无重复试验的多因素试验设计)、可考察部分交互作用的多因素设计(有重复试验的随机区组设计、有重复试验的平衡不完全随机区组设计、有重复试验的拉丁方设计、交叉设计、裂区设计、具有重复测量的多因素设计、含区组因素的析因设计、正交设计、稳健设计)和可考察全部交互作用的多因素设计(如析因设计)。

在回归多因素设计中,常分为均匀设计、混料设计(如单型格子混料设计、单型重心混料设计、有下界约束混料设计、极端顶点混料设计、混料均匀设计、乘积混料设计等)、饱和设计与超饱和设计、各种最优设计(A最优设计、D最优设计、G最优设计、L最优设计)和各种回归设计(如二次正交旋转组合设计、二次通用旋转组合设计等)。

在临床试验中,常需要对两种处理(如两种药物或一种药物的两个剂量)的疗效或安全性进行比较或假设检验,传统的做法是两种处理下的主要评价指标数值(均值或率)相减求得差量,看此差量是否接近于 0。也就是说,检验“差量”与“0”之间的差别是否具有统计学意义,这样的比较思想和做法被称为“差异性检验”。然而,在临床实践中,人们往往并不满足基于上述思想所得出的结论。于是,随着人们对事物认识的程度不断深入,就逐渐产生出另外三种特殊的比较,它们被称为“非劣效性检验”、“等效性检验”和“优效性检验”。这三种特殊检验被称为新的“比较类型”。事实上,这三种特殊的比较类型在本质上仍属于“差异性检验”。所不同的是,三种特殊比较类型强调的是:两对比组主要评价指标之“差量”与一个非 0 的且在临床上有意义的“非 0 界值 δ ”项比较。

在非劣效性检验中,非 0 界值被称为“非劣效性界值,常记为 δ_L ,而且, δ_L 常取负值,“负号”仅表示方向”;在优效性检验中,非 0 界值被称为“优效性界值,常记为 δ_U ,而且, δ_U 常取正值”;在等效性检验中,非 0 界值由两个“等效界值 δ_L 和 δ_U ”分别取代。三种特殊比较的最终结果与各自相应的“非 0 界值 δ ”关系很大!通常需要由多位来自不同地区且具有相近学术知名度的临床专家共同讨论后商定。四种比较类型的含义如下。

(1)差异性检验:差异性检验是指主要研究目的为显示两种治疗效果之差量与 0 之间的差别是否具有统计学意义的试验。在试验设计阶段不需要设定任何界值。具体实施检验时可分为单侧检验(左单侧检验或右单侧检验)与双侧检验。

(2)非劣效性检验:非劣效性检验是指主要研究目的为显示试验药的治疗效果在临床上不比阳性对照药差的试验。在试验设计阶段需要设定一个界值 δ_L ,来界定试验药是否不比阳性对照药疗效差过预先设定的这个界值,它是一个单侧检验。

(3)等效性检验:等效性检验是指主要研究目的为显示两种治疗效果之间的差别大小在临床上并无重要意义的试验。在试验设计阶段需要设定等效性界值(δ_L, δ_U)来界定两种治疗的等效性,它是双单侧检验,即在左、右两端各进行一次单侧检验。

(4)优效性检验:优效性检验是指主要研究目的为显示试验药的治疗效果优于对照药(安慰剂对照或阳性对照)的试验。在试验设计阶段需要设定一个界值 δ_U ,来界定试验药的优效性,它是一个单侧检验。

在成组设计下,如何实施四种比较类型下的样本含量与检验效能估计、如何进行四种比较类型下的一元定量和定性资料的假设检验等内容,可参阅有关统计学专著,此处从略。

19.2.8 实时并严格控制临床试验研究质量

一个科研课题的设计质量对于其最终结果是至关重要的,但课题必须付诸实施才能产生结果(数据或产物或产品)。由于课题实施是一个较长、甚至漫长的过程,在这个过程中,有无穷多个(其中有一些是十分重要的)非试验因素随时“渗透”进来,研究结果很容易被这些非试验因素所沾染,甚至是嵌入式的“腐蚀”科研结果。因此,实时并严格控制临床试验研究质量就显得十分重要且必不可少!

质量控制的结局是应确保科研数据准确可靠。因此,质量控制的源头是“找准找全重要的非试验因素”和“采取有效措施控制非试验因素对结果的影响”。重要非试验因素来自何处?主要来自以下四个方面:①控制中心效应;②控制研究者效应;③控制受试者效应;④控制环境和条件效应。控制的具体内容和做法,概述在图 19-1 和 19-2 中。

值得一提的是:在临床试验研究过程中,质量控制做得好坏,最突出的标志是看“中心效应”是否明显。曾经有过这样一个临床研究项目,其最终的中心效应十分明显:表现在某主要疗效评价指标在不同中心之间的差别具有统计学意义(其最低值为 22.3%、最高值为 78.2)。严格地说,此项临床试验应视为失败了!



图 19-1 重要非试验因素来源与质量控制要领

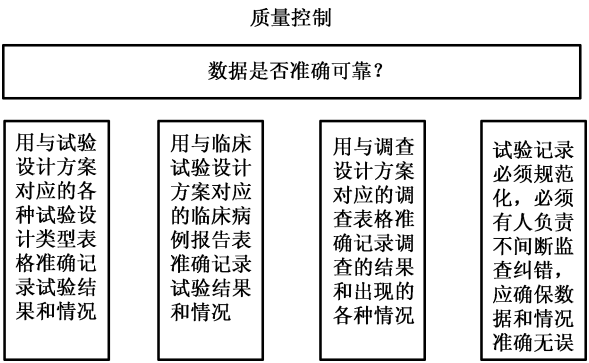


图 19-2 消除或降低重要非试验因素影响的具体途径

由前面的论述可知，导致一项临床试验出现明显的中心效应的原因虽然很多，但最主要来自研究者、受试者和环境及条件。特别是各临床试验中心所用的仪器设备、试剂、测定方法、指标的单位等在质量、型号、批次、定义、技术方法等各方面可能存在实质性的差别。如果条件允许，应尽可能设置“中心试验室”。来自研究者方面的误差影响，主要是靠统一培训和严格考试来降低；而来自受试者和研究者心理因素方面的误差影响，主要是靠真正意义上的双盲或三盲措施来降低。

19.2.9 正确无误地在临床病例报告表上进行记录

在临床试验研究中，首先需要制订出科学完善的临床试验设计方案，然后，在此方案指导下，编制出条理清楚、内容全面的临床病例报告表(简称 CRF)。临床试验研究的最终结果是否准确可靠取决于每一份 CRF 填写得是否准确。所以，全部参与临床试验研究的工作者在正式参与临床试验之前，必须接受统一的思想教育(思想上应高度重视、应具有很强的责任心和使命感)和技术培训(应精通业务、技术熟练)，以便确保在整个临床试验过程中将来自研究者的各种非试验因素的影响效应降至最低程度，为收集值得进行统计分析的临床试验数据打下坚实的基础。

(胡良平 胡纯严 陶丽新 李子建 刘惠刚)

第 20 章 临床科研资料统计表达描述要览

在临床科研资料表达描述中，必须牢牢抓住以下几个核心内容：如何正确地运用统计表、如何以简化形式呈现定性资料、如何以简化形式呈现多组定量资料、如何以数据库的形式呈现多因素多指标的复杂临床科研资料、如何合理地选用统计图。

20.1 如何正确地运用统计表

20.1.1 常见统计表的种类

统计表常可分为：简单统计表(罗列各种现象或情况，其各组数据的合计值无意义)、分组统计表(全部受试者或样品或数据按某一共同的变量或属性被分成多个组，各组之间的互不交叉、也不重叠，全部各组受试者的合计等于总样本含量)、复合统计表(全部受试者或样品或数据同时按两个或多个变量被分组，全部小组数等于各分组变量的水平数之乘积)。

20.1.2 常见统计表举例

1. 简单统计表

【例 20-1】 在表 20-1 中，列出了 12 种心理问题，原作者根据各行数据与总患者人数 50 计算出 12 个百分率。

【说明】 之所以说此表格属于简单统计表，是因为每位受试者可以有多种心理问题，全部被调查的受试者只有 50 名，但出现 12 种心理问题的人数之和明显地超出了此总数，因此，简单表中绝对数的合计值是无意义的。

2. 分组统计表

(1) 具有一个定性原因变量和一个定性结果变量的分组表。

【例 20-2】 在表 20-2 中，收集了 5 种药物治疗 180 例精神抑郁症患者所得疗效的记录结果。

表 20-1 病人术前后常见的心理问题(N=50)

心理问题	人数	百分率(%)
恐惧手术	41	82.0
希望有一个老医生	40	80.0
希望术后病痛解除	34	68.0
希望手术细心	30	60.0
担心为恶性病变	28	56.0
希望了解手术方案	26	52.0
希望医生多关心	34	68.0
希望亲人陪护	31	62.0
希望护士做好出院指导	29	58.0
担心远期并发症	25	50.0
担心手术未成功	17	34.0
担心丧失工作能力	11	22.0

表 20-2 5 种药物治疗精神抑郁症的疗效

药物种类	例数		
	疗效：改善	未改善	合计
安慰剂	8	22	30
A 药	12	18	30
B 药	21	9	30
C 药	15	15	30
D 药	14	16	30
E 药	19	11	30
合计	89	91	180

【说明】之所以说此表格属于分组统计表，是因为每位受试者只能接受一种药物治疗，6个药物组中的受试者数目之和等于总样本含量。由于该表中的分组变量是一个离散型变量(药物种类)，故此表在统计学上有另一个专业名称，称为二维[一个“维度”指原因变量(即药物种类)、另一个“维度”指结果变量(即疗效)，其实，就是有两个变量]列联表资料。

(2)具有一个定量结果变量的分组表。

【例 20-3】某地 150 名正常成年男子红细胞数($10^{12}/L$)的资料见表 20-3。

表 20-3 某地 150 名正常成年男子红细胞数($10^{12}/L$)频数分布表

组段/($10^{12}/L$) (1)	频数 f (2)	组中值 X_0 (3)	$f X_0$ (4) = (2) \times (3)
3.7 ~	1	3.8	3.8
3.9 ~	4	4.0	16.0
4.1 ~	11	4.2	46.2
4.3 ~	17	4.4	74.8
4.5 ~	26	4.6	119.6
4.7 ~	32	4.8	153.6
4.9 ~	26	5.0	130.0
5.1 ~	18	5.2	93.6
5.3 ~	10	5.4	54.0
5.5 ~	4	5.6	22.4
5.7 ~5.9	1	5.8	5.8
合 计	150	—	719.8

【说明】之所以说此表格属于分组统计表，是因为每位受试者的测定值只能落在某一个组段中，各组段上的人数之和就等于总样本含量。由于该表中的分组变量是一个连续型变量[红细胞数($10^{12}/L$)]，故此表在统计学上有另一个专业名称，叫做定量资料频数分布表。

3. 复合统计表

(1)具有两个定性原因变量和一个定性结果变量的复合表。

【例 20-4】某临床医生收集到如表 20-4 所示的资料，该资料反映了孕妇在两个诊所接受的产前护理量与婴儿存活情况的观察结果。

表 20-4 孕妇在两个诊所接受的产前护理量与婴儿存活情况的观察结果

接受护理地点	产前护理量	婴儿例数		死亡率(%)	
		婴儿存活状况	死		活
诊所 A	少		3	176	1.676
	多		4	293	1.347
诊所 B	少		17	197	7.944
	多		2	23	8.000

【说明】在此表中，接受护理地点和产前护理量是两个定性的原因变量，同时，它们也是“试验分组变量”，它们共形成 4 种不同的水平组合，将全部孕妇分成 4 个小组，故可称其为“具有两个定性原因变量和一个定性结果变量的复合表”；此表在统计学上常称为“结果变量为二值变量的三维列联表”。

(2)具有三个定性原因变量和一个定性结果变量的复合表。

【例 20-5】某研究者收集到如表 20-5 所示的资料，该资料反映了年龄、性别和是否吸烟对是否患肺癌的影响。

表 20-5 同时按年龄和性别分层后吸烟与否与是否患肺癌的调查结果

年龄(岁)	性别	吸烟与否	例数		
			患肺癌与否	患	未患
≤40	男	吸烟		5	5
		不吸烟		60	55
	女	吸烟		40	50
		不吸烟		5	5
>40	男	吸烟		30	10
		不吸烟		30	5
	女	吸烟		5	55
		不吸烟		5	35

【说明】 在此表中，年龄分组、性别和吸烟与否是三个定性的原因变量，同时，它们也是“试验分组变量”，它们共形成 8 种不同的水平组合，将全部受试者分成 8 个小组，故可称其为“具有三个定性原因变量和一个定性结果变量的复合表”；此表在统计学上常称为“结果变量为二值变量的四维列联表”。

值得注意的是，此表中的数据可能不是真实的调查结果，因为有两点十分可疑：第一，所有数据的个位数不是“0”就是“5”；第二，第 1 行和第 4 行上对应着两个不同的条件，各条件下都只调查了 10 位受试对象，居然有一半患了肺癌。在如此小的样本含量的样本中，患肺癌的概率如此之高，不符合实际，这种情况一般只有在医院“肺癌病房内”才有可能出现。因为在某个大肺癌病房中住了 5 位肺癌患者，还有 5 位未患肺癌的“护工或家属”。由以上两点猜疑可推知，此资料中的数据很可能是原文作者(Simpson, 1951)杜撰出来的，引用时务必加以说明。

(3)具有四个定性原因变量和一个定性结果变量的复合表。

【例 20-6】 某研究者收集到如表 20-6 所示的资料，该资料反映了乳腺癌患者被确诊的医疗中心、年龄、慢性炎症反应程度、核的量级对患者 3 年后的结局(死与活)的影响情况。

表 20-6 4 个因素影响下乳腺癌患者三年存活情况的调查结果

患者被确诊的 医疗中心	患者年龄(岁)	慢性炎症 反应的 程度	核的量级	例数		
				三年结局	死	活
东京	<50	轻	恶性		9	26
			良性		7	68
		重	恶性		4	25
			良性		3	9
	50 ~	轻	恶性		9	20
			良性		9	46
		重	恶性		11	18
			良性		2	5
	70 ~	轻	恶性		2	1
			良性		3	6
		重	恶性		1	5
			良性		0	1
波士顿	<50	轻	恶性		6	11
			良性		7	24
		重	恶性		6	4
			良性		0	0

续表						
患者被确诊的 医疗中心	患者年龄(岁)	慢性炎症 反应的程度	核的量级	例数		
				三年结局	死	活
格拉摩根	50 ~	轻	恶性		8	18
			良性		20	58
		重	恶性		3	10
			良性		2	3
	70 ~	轻	恶性		9	15
			良性		18	26
		重	恶性		3	1
			良性		0	1
	<50	轻	恶性		16	16
			良性		7	20
		重	恶性		3	8
			良性		0	1
	50 ~	轻	恶性		14	27
			良性		12	39
		重	恶性		3	10
			良性		0	4
	70 ~	轻	恶性		3	12
			良性		7	11
		重	恶性		3	4
			良性		0	1

【说明】 在此表中，乳腺癌患者被确诊的医疗中心、年龄、慢性炎症反应程度、核的量级是四个定性的原因变量，同时，它们也是“试验分组变量”，它们共形成 $3 \times 3 \times 2 \times 2 = 36$ 种不同的水平组合，将全部受试者分成 36 个小组，故可称其为“具有四个定性原因变量和一个定性结果变量的复合表”；此表在统计学上常称为“结果变量为二值变量的五维列联表”。

(4)具有两个定性原因变量和一个定量结果变量的复合表。

【例 20-7】 某研究者收集到如表 20-7 所示的资料，该资料反映了骨髓红系祖细胞分别在两种药物不同的浓度作用下，骨髓红系祖细胞的增殖情况。

表 20-7 PDS 与 PTS 对骨髓红系祖细胞的增殖作用($\bar{x} \pm s, n = 10$)								
药物种类	骨髓红系祖细胞的含量(个/ 0.5×10^5)							
	药物浓度 (ng/mL) :	0.0	2.5	12.5	25.0	50.0	100.0	200.0
PDS		115.2 \pm 20.6	143.8 \pm 20.8	144.3 \pm 29.6	154.3 \pm 28.2	178.5 \pm 30.2	165.9 \pm 22.4	152.5 \pm 25.2
PTS		115.2 \pm 20.6	95.9 \pm 12.9	89.1 \pm 14.0	85.8 \pm 14.3	49.5 \pm 12.9	6.5 \pm 2.7	—

注：PDS 为人参二醇皂苷；PTS 为人参三醇皂苷。

【说明】 在此表中，受试对象是“骨髓红系祖细胞”。药物种类是一个二值定性原因变量，药物浓度是一个多值有序定性原因变量，而骨髓红系祖细胞的含量是一个定量的结果变量。

(5)具有三个定性原因变量和两个定量结果变量的复合表。

【例 20-8】 某研究者收集到如表 20-8 所示的资料，该资料反映了中西医的诊断分类条件下所对应的两个定量指标的取值大小。

表 20-8 各组 NSCLC 肺部 ROI 的 MaxSUV 和 MeanSUV 比较($\bar{x} \pm s$)

是否为血瘀证	癌症类型	癌症分期	例数	MaxSUV	MeanSUV
非血瘀证	腺癌	I	X	X	X
		II	X	X	X
		III	X	X	X
		IV	X	X	X
	鳞癌	I	X	X	X
		II	X	X	X
		III	X	X	X
		IV	X	X	X
	大细胞癌	I	X	X	X
		II	X	X	X
		III	X	X	X
		IV	X	X	X
	血瘀证	腺癌	I	X	X
			II	X	X
			III	X	X
			IV	X	X
		鳞癌	I	X	X
			II	X	X
			III	X	X
			IV	X	X
	大细胞癌	I	X	X	X
		II	X	X	X
		III	X	X	X
		IV	X	X	X

注：表中最后两列用“X”代替对应条件下的“ $\bar{x} \pm s$ ”；而倒数第 3 列用“X”代表对应条件下的样本含量。因原文作者错误地将一张表拆分成了两张表,故数据无法还原。

【说明】 在此表中，受试对象是三种癌症患者。此资料先将全部患者按中医分成两大类(即非血瘀证与血瘀证)，在中医分类中的每一类，再按西医的两种分类方法(癌症类型与癌症分期)划分，它们共形成 $2 \times 3 \times 4 = 24$ 种不同的水平组合，将全部受试者分成 24 个小组，故可称其为“具有三个定性原因变量和一个定量结果变量的复合表”。

对于如何以简化形式呈现定性资料(见表 20-4 ~ 表 20-6)、如何以简化形式呈现多组定量资料(表 20-7 和表 20-8)，此处不再赘述。

20.2 如何以数据库的形式呈现多因素多指标的复杂临床科研资料

20.2.1 概述

虽然可以采用 20.1 节介绍的常用统计表来呈现多因素多指标的科研资料，但当因素和指标都很多时，显然是不太合适且很不方便的。例如，在某个复杂临床试验研究项目中，有 100 个原因变量、有 100 个结果变量，如何编制“复合统计表”呢？若将全部信息列在一张复合统计表中，此表无法被正确地呈现出来；若将全部信息进行拆分，采用多张复合统计表来呈现，第 1 个问题是如何拆分(拆分结果因人而异)？第 2 个问题是破坏了原先的整体设计、割裂了变量之间原本存在的数量联系。为了避免直接处理上述两个棘手的问题，需要另辟蹊径。

20.2.2 以数据库的形式呈现多因素多指标的复杂临床科研资料

所谓以数据库的形式呈现多因素多指标的复杂临床科研资料，其实很简单，就是把每位受试者的全部信息按相同的顺序写成一行，每列代表一个变量及其在每个个体身上的取值或表现。例如，在下面的例子及表 20-9 中，前面 103 行与后面 100 行信息分别对应着 103 位冠心病患者与 100 位正常对照者；而各列的含义由表头不言自明，第 1 列“编号”仅表明各受试者的相对位置，对于资料的统计分析没有什么实际意义，其他各列都有具体的专业含义。根据不同的统计分析目的，需要从第 2 列到最后一列中，抽取不同的列来进行分析。当然，也可先从 203 行中先抽取不同的行（比如，分别抽取前 103 行或后 100 行），再从第 2 列到最后一列中，抽取不同的列来进行分析。各种各样的抽取方法对应着各种简单到复杂的统计分析方法，全部都系统地实现了，就能写出一本很厚的统计学书。

【例 20-9】 有人对 103 例冠心病患者 ($G=1$) 和 100 例正常对照者 ($G=2$) 进行了多项指标的观测，资料见表 20-9。

表 20-9 冠心病人与正常人多项指标的观测结果

编号	组别	性别	年龄	高血压史	吸烟史	胆固醇	甘油三酯	低密度脂蛋白	高密度脂蛋白	脂蛋白 α	载脂蛋白 As	载脂蛋白 B	基因型 xbal	基因型 EcoRI	服药情况
N	G	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄
1	1	男	60	无	无	223	205	122	30	106	0.92	0.74	-/-	-/-	未服
2	1	女	46	无	无	166	51	84	57	56	1.14	0.54	-/-	+/-	β 阻滞剂
3	1	男	55	有	无	273	155	197	34	58	1.01	0.93	+/-	+/+	β 阻滞剂
4	1	男	30	无	有	193	99	70	37	85	0.94	0.70	-/-	+/+	降血脂药
...
102	1	女	66	无	无	194	77	107	54	43	0.98	0.61	-/-	+/+	未服
103	1	男	76	有	有	195	101	135	42	302	0.91	0.65	-/-	+/+	β 阻滞剂
104	2	女	39	无	无	223	48	157	55	32	1.06	0.71	-/-	+/+	未服
105	2	男	40	无	无	152	91	100	46	62	1.00	0.55	-/-	-/-	未服
106	2	男	59	无	有	168	84	106	33	205	0.67	0.69	-/-	+/+	未服
...
203	2	男	69	有	无	224	110	58	49	132	1.10	0.96	-/-	+/+	未服

【说明】 在此表中，“组别、性别、年龄、高血压史、吸烟史、基因型”通常可被视为影响因素，而血脂指标 $X_5 \sim X_{11}$ 可被视为定量的结果变量。但这并非是一成不变的，人们也可将“是否患冠心病”视为“二值结果变量”，而把血脂指标 $X_5 \sim X_{11}$ 视为定量的原因变量。

【例 20-10】 有人从某地区随机地抽取了 50 人，资料见表 20-10。

表 20-10 从某地随机抽取的 50 人有关指标的测定结果

NO	Race	Sex	Age	Education	Weight	Height	Smoke	SBP	BMI
1	1	1	28	16	68	160	7	111	24.33
2	1	1	26	12	68	165	1	101	25.09
3	2	2	31	15	68	175	1	120	26.61
4	2	1	18	12	76	265	7	158	32.26
5	1	1	50	17	67	145	1	125	22.71
6	2	1	42	12	69	247	1	166	36.48
7	1	2	20	12	66	156	7	114	25.18

									续表
NO	Race	Sex	Age	Education	Weight	Height	Smoke	SBP	BMI
8	1	1	29	12	76	180	1	143	21.91
9	1	2	35	12	63	166	2	111	29.41
10	1	1	47	16	66	169	1	133	27.28
11	1	2	20	14	69	120	7	95	17.72
12	1	2	33	16	68	133	7	113	20.22
13	4	1	24	13	71	185	7	128	25.80
14	1	1	28	14	72	150	1	110	20.34
15	1	2	32	8	61	126	1	117	23.81
16	2	1	21	10	68	190	1	112	28.89
17	1	1	28	17	71	150	7	110	20.92
18	1	2	60	12	61	130	7	117	24.56
19	1	1	55	12	66	215	2	142	34.70
20	1	2	74	12	65	130	7	105	21.63
21	1	2	38	16	68	126	7	94	19.16
22	1	1	26	14	66	160	2	131	25.82
23	1	1	52	9	74	328	2	128	42.11
24	1	2	25	16	69	125	7	93	18.46
25	1	2	24	12	67	133	1	103	20.83
26	1	2	26	16	59	105	1	114	21.21
27	1	2	51	13	64	119	7	130	20.43
28	2	2	29	16	62	98	7	105	17.92
29	4	1	26	0	64	150	7	117	25.75
30	1	2	60	12	64	175	1	124	30.04
31	1	1	22	9	70	190	1	122	27.26
32	1	2	19	12	65	125	7	112	20.80
33	3	1	39	12	73	210	1	135	27.71
34	3	2	77	4	62	138	7	150	25.24
35	1	1	39	12	73	230	2	125	30.34
36	1	1	40	11	69	170	1	126	25.10
37	1	2	44	13	62	115	7	99	21.03
38	3	2	27	9	61	140	7	114	26.45
39	1	1	29	14	73	220	7	139	29.03
40	1	2	78	11	63	110	7	150	19.49
41	1	1	62	13	65	208	7	112	34.61
42	1	1	22	10	71	125	1	127	17.43
43	1	2	37	11	64	176	7	125	30.21
44	1	1	38	17	72	195	7	136	26.45
45	3	1	22	12	65	140	7	108	23.30
46	3	1	79	0	61	125	2	156	23.62
47	1	2	24	12	62	146	7	108	26.70
48	1	2	32	13	67	141	2	105	22.08
49	1	1	42	16	70	192	7	121	27.55
50	1	1	42	14	68	185	7	126	28.13

【说明】 在此表中，第 1 列“NO”代表编号、第 2 列“Race”代表人种、第 3 列“Sex”代表性别、第 4 列“Age”代表年龄、第 5 列“Education”代表接受教育的年数、第 6 列“Weight”代表体重(kg)、第 7 列“Height”代表身高(cm)、第 8 列“Smoke”代表三种抽烟量、第 9 列“SBP”代表收缩压(mm-Hg)、第 10 列“BMI”代表体重指数。

20.3 如何合理地选用统计图

20.3.1 概述

为了形象化地呈现科研资料，人们经常绘制统计图。统计图一般分为三大类：第一类为表达定性资料(强调结果变量为定性的变量)的统计图、第二类为表达定量资料(强调结果变量为定量的变量)的统计图、第三类为同时反映两个或三个定量变量变化趋势的统计图。然而，面对拟表达的科研资料，人们常错误地选用统计图。最常见的错误有以下几种。

第一，误用表达离散型或称定性资料的统计图表达连续型或称定量资料；第二，坐标轴上所标的刻度值违反了数学原则(三种错误的情形：其一，横坐标轴上从左到右数值由大到小，纵坐标轴上由下向上数值由大到小；其二，直角坐标系的原点不是(0, 0)点；其三，同一个坐标轴上，以普通算术尺度表达，但等长的“间隔”代表的“数量”却不等，如 3, 5, 8, 14, 18, …)。由此可知，如何合理地选用统计图呈现科研资料是一个必须明确回答的问题。

20.3.2 合理选择统计图表达定性资料

1. 如何呈现无约束的多值名义变量的频数分布

【例 20-11】 沿用例 20-1 的资料，试选用合适的统计图类型将其表达出来。

【分析与解答】 这个资料很容易误认为是“计数资料”，因为出现不同心理问题的有几个人。事实上，当研究者考察每一位受试者时，记录的是每种心理问题是“出现”还是“没有出现”，这显然是“二分类资料”，即“二值变量的结果”。对每一种心理问题而言，对每一位受试者进行观察，相当于做一次试验，其结果在统计学上被称为一个“两点分布问题”，对全部 50 个受试者做完 50 次试验，相当于把 50 个“两点分布结果”加在一起，在统计学上就称为一个“二项分布问题”。因此，出现某种心理问题就是一个“定性结果”，12 种心理问题就相当于一个定性结果变量有 12 个不同的取值或表现，而标志每种“取值或表现”出现频繁程度的数据就是“频数或人数”。由于这 12 种心理问题是“多值名义变量”，它们在横坐标轴上被放置的位置可以是任意的，故反映各种心理问题出现的频数大小的统计图形适合选用描述多值名义变量的“单式条图”(如图 20-1 所示)。

2 如何呈现有约束的多值名义变量的频数分布

【例 20-12】 某医院 1990 年和 1998 年住院病人死于五种疾病的人数见表 20-11，试选用合适的统计图将其表示出来。

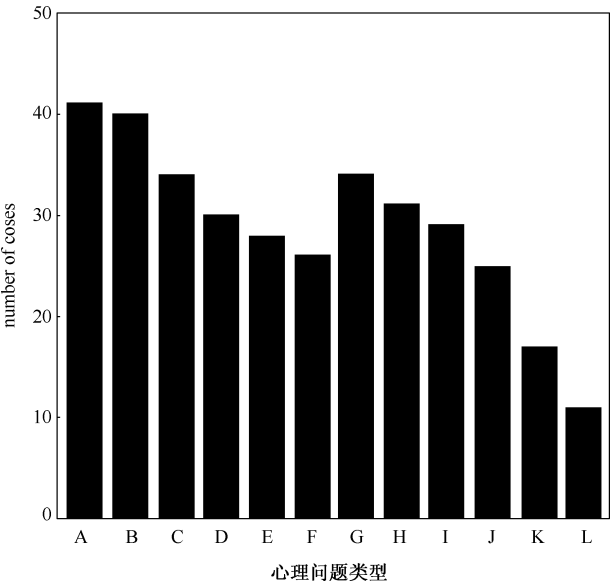
表 20-11 某医院 1990 年和 1998 年住院病人的五种疾病死亡人数和构成比

疾病构成	死亡人数		构成比(%)	
	1990 年	1998 年	1990 年	1998 年
恶性肿瘤	58	40	30.53	26.85
循环系统疾病	44	44	23.16	29.53
呼吸系统疾病	37	29	19.47	19.46

续表

疾病构成	死亡人数		构成比 (%)	
	1990 年	1998 年	1990 年	1998 年
消化系统疾病	19	18	10.00	12.08
传染病	32	18	16.84	12.08
合计	190	149	100.00	100.00

【分析与解答】 这是描述分类变量的各类别所占的构成比，同时还存在两个不同年份构成情况的对比，有两种可选用的统计图，分别称为百分条图和复式圆图，如图 20-2 和图 20-3 所示。所谓有约束，指的是各组相对数之和为 100%。



A: 恐惧手术; B: 希望有一个老医生; C: 希望术后病痛解除; D: 希望手术细节;
E: 担心为恶性病变; F: 希望了解手术方案; G: 希望医生多关心; H: 希望亲人陪护;
I: 希望护士做好出院指导; J: 担心远期并发症; K: 担心手术未成功; L: 担心丧失工作能力

图 20-1 病人术前术后常见的心理问题出现频数单式条图

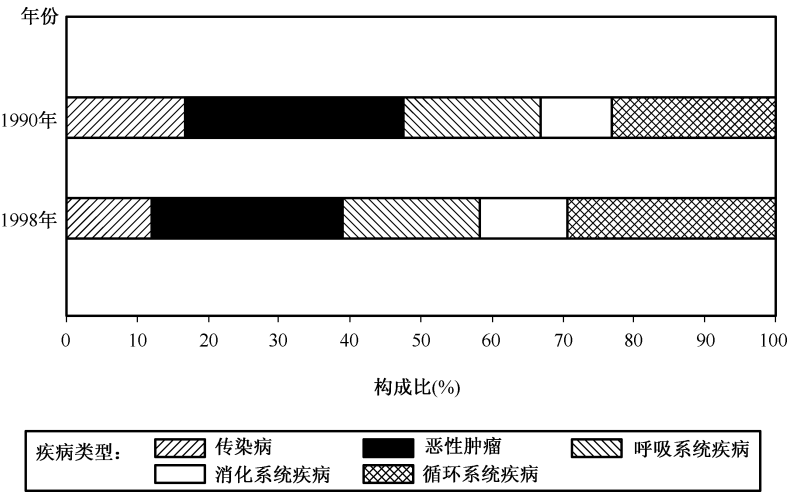


图 20-2 某医院 1990 年和 1998 年住院病人的五种疾病死亡构成比百分条图

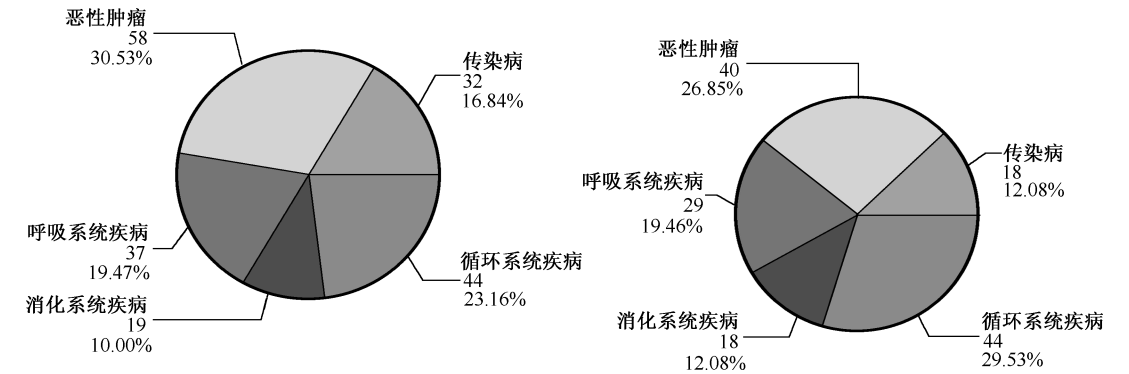


图 20-3 某医院 1990 年和 1998 年住院病人的五种疾病死亡构成比复式圆图

3. 如何呈现单组设计一元定量资料的频数分布

【例 20-13】 沿用例 20-3 的资料，试选用合适的统计图呈现其频数分布情况（如图 20-4 所示）。

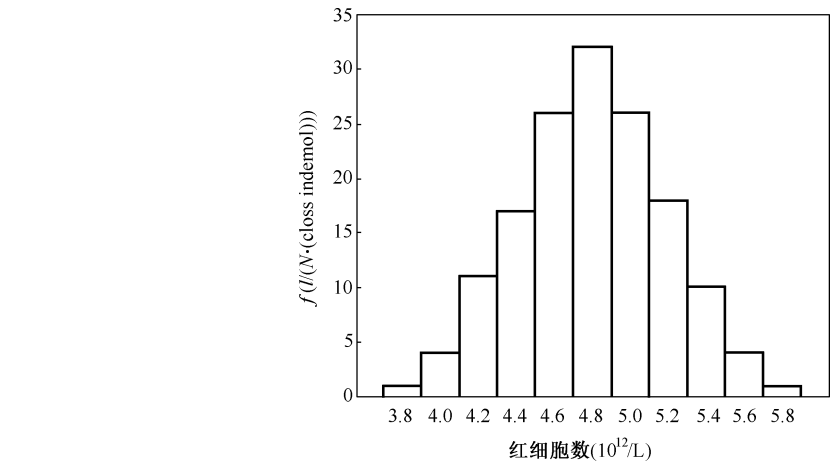


图 20-4 某地 150 名正常成年男子红细胞数 (10¹²/L) 的频数分布直方图

4. 如何用图式法直观判断单组设计一元定量资料是否服从某特定分布

在表 20-10 中，年龄 (age)、体重 (weight)、身高 (height)、收缩压 (SBP)、体重指数 (BMI) 这 5 个定量变量都可以被视为来自单组设计的定量资料。若希望通过形象化的手法来判断其中每一个定量变量的频数分布规律是否服从正态分布或指数分布或威布尔分布，可绘制 P - P 图或 Q - Q 图或概率图，其横坐标轴上通常放置假定的某种分布对应的理论累积频率 (对 P - P 图而言) 或理论分位数 (对 Q - Q 图和概率图而言)，而纵坐标轴上通常放置实际累积频率 (对 P - P 图而言) 或实际分位数 (对 Q - Q 图而言) 或实际观测值 (对概率图而言)。此时，若图中的全部散点在图中对角线上下附件随机地散布着 (最好完全落在这条直线上)，就可以认为所描述的单组设计一元定量资料近似服从特定的那种分布规律。具体地说，三种图的绘制原理如下。

(1) P - P 图：将定量指标的观测值按由小到大的顺序排列，然后计算其实际累积频率，以实际累积频率对被检验分布 (例如正态分布) 的理论累积频率作图，称为 P - P 图。若原资料确实服从假定的概率分布，那么，图形大体上将呈直线。

(2) Q - Q 图: 将定量指标的观测值按由小到大的顺序排列, 以实际观测值对被检验分布 (例如正态分布) 的理论分位数作图, 称为 Q - Q 图。若原资料确实服从假定的概率分布, 那么, 图形大体上将呈直线。

(3) 概率图: 将定量指标的观测值按由小到大的顺序排列, 以实际观测值对被检验分布 (例如正态分布) 的理论百分位数作图, 称为概率图。若原资料确实服从假定的概率分布, 那么, 图形大体上将呈直线。

具体做法和绘出的图形, 此处从略。

5. 如何呈现单组设计二元定量资料的变化趋势

(1) 散点图或散布图。

适合描述两个在专业上有一定联系且属性比较接近的定量变量的同时变化趋势, 如描述一组正常成年人身高与体重的同时变化趋势、血浆清蛋白含量与血红蛋白含量的同时变化趋势等。

(2) 普通线图与半对数线图。

适合用来描述时间序列数据 (如各年份某地区某种疾病死亡率数据) 随时间的推移的动态变化趋势。普通线图适合在一幅图中描述一个时间序列数据 (例如, 仅一种病的死亡率) 随时间推移的动态变化趋势; 而半对数线图适合在一幅图中描述两个或多个时间序列数据 (例如, 有多种类似疾病的死亡率) 随时间推移的动态变化趋势。

(3) 误差线图。

在适合绘制普通线图的场合下, 若每个时间点上做了 $k(k \geq 2)$ 次独立重复试验, 设因变量为 y , 自变量为 t , 则可用数据点 $(t_i, \bar{y}_i) (i = 1, 2, \dots, n)$ 来绘制普通线图, 但在每个试验点上因变量 y 都有 k 个不同的数值, 当然可以计算出标准差, 将各点上因变量 y 的标准差用类似“天线”的形式标示出来, 这样得到的线图就被称为带误差的线图, 如图 20-5 所示。

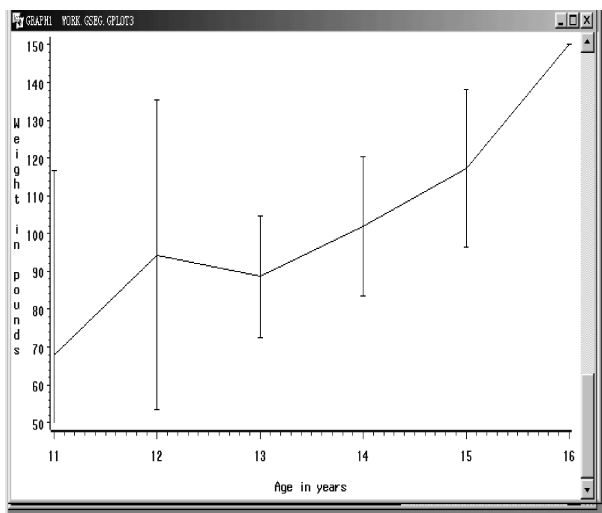


图 20-5 反映年龄与体重之间变化趋势的误差线图

【说明】 适合绘制线图的资料中, 放置在横坐标轴上的变量不一定必须是“时间”, 原则上只要是定量变量都可以。在图 20-5 中, 年龄 (age, 岁) 就是一个普通的定量变量, 只有相同年龄的人有多个, 他们的体重就可能取多个不同的数值, 因此, 与每个年龄对应的多个体重数值就可算出一个算术平均值和标准差, 于是, 就可绘制出带误差的线图。

(胡良平 胡 完 郭春雪 李 双 刘一松)

第 21 章 临床科研资料统计分析要览

临床科研过程之中和结束之后会产生大量的科研资料，如何表达和描述这些科研资料的方法在本书第 20 章做了比较详细的介绍。本章重点讨论如何基于统计表达与描述，对科研资料进行探索性统计分析和正式的统计分析。众所周知，正确地运用统计学，统计分析就是揭示事物或现象内在变化规律的强有力的思想方法和技术手段；而错误地运用统计学，统计分析就会变成帮人造假的“魔咒”。

21.1 统计分析方法的分类

21.1.1 概述

统计分析的前提是“数据值得分析”。如何确保拟分析的数据是值得分析的呢？关键是要有科学完善的科研设计方案的指导和实时且严格的质量控制，由此而获得的准确可靠且原始的科研资料（任何数据管理和预处理都没有改变数据的本质及其原始性）才是值得进行统计分析的。

当今社会，很多级别各异的单位都在热衷于建立“数据中心”。投入的科研经费不计其数，但由于他们缺乏正确的统计思想和统计设计理念及知识，致使一些所谓的“数据中心”成了堆积杂乱无章数据的“仓库”。这些数据除了要花费大量的人力物力去保存之外，很难发挥其应有的作用。十几年前，全人类都在热衷于炒作“数据挖掘”；而现如今，全世界又正在狂热地炒作“大数据”。可以预见，这种忽视统计思想和统计知识指导而一味地盲目炒作是没有什么实际意义的，也是没有什么发展前途的。

对值得分析的数据如何选择合适的统计分析方法予以处理，这是所有需要处理数据的人们不可回避的一个问题。要想做到正确选择统计分析方法处理给定的资料，首先必须清楚统计分析方法的分类。

迄今为止，全人类创造出来的统计分析方法不计其数。事实上，由于命名的规则或角度不同而导致了有些统计分析方法互相交叉或重叠。例如，经典统计学派与贝叶斯统计学派分别有各自的命名规则和结果。即使在同一种学派内部，还可根据同时分析的结果变量的性质与个数、原因变量的个数、统计分析所能实现的功能或目标等的不同，又有许多不同的命名规则和结果。

从把“估计量”视为“固定参数”还是“随机变量”角度来划分，统计分析方法可分为两大类，每一大类又可分为四个中类，每一个中类再分成若干小类。两大类分别为“一水平统计分析方法”与“多水平统计分析方法”。它们之间的区别就在于全部拟被分析的数据是仅来自单一层次，还是来自两层或多个层次。而每一大类统计分析方法中又都可以分成四个中类，即探索性统计分析、差异性统计分析、相关与回归分析、广义综合评价。值得注意的是，在前面提及的四个中类的每一类中，还时常会伴随运用“区间估计”的统计分析方法，因为此法是对前已提及的主要功能统计分析方法的补充和完善。

有人可能会问：为何没有包括“多因素统计分析”和“多元统计分析”？那是因为这两种说法都是非常含糊的，“多因素统计分析”强调的是考察的原因变量的个数大于等于 2 的所有统计分析方法，它们通常包括多因素差异性分析方法和一元多重回归分析方法；而“多元统计分析”则强调的是每次分析的结果变量的个数大于等于 2，但也有一些统计学家又常把“一个因变量多个自变量的

统计分析方法”稀里糊涂地装进了多元统计分析的“口袋”中。事实上,从功能上来划分,现有的多元统计分析方法中至少包括了三类统计分析方法,即差异性分析方法、相关与回归分析方法和广义综合评价方法。

21.1.2 探索性统计分析

什么是探索性统计分析?就是了解资料的基本情况的描述性统计分析,通常是通过编制统计表、绘制统计图、用简化方法呈现科研资料中各定量变量的平均水平和离散程度;探索性统计分析的效果是:第一,呈现单组设计一元定量或定性变量的频数分布规律;第二,同时呈现两个或多个定量变量同时变化的趋势。探索性统计分析的目的是:为拟进行的正式的某一类统计分析“探路”,即了解资料是否满足某些参数检验的前提条件、寻找合适的变量变换方法、为是否需要引入某种形式的派生变量提供线索、产生哑变量前如何选定基准水平等。

21.1.3 差异性统计分析

差异性统计分析是指两个或多个同类统计量之间的差量与 0 之间的差别是否具有统计学意义的一大类统计分析方法的统称。有些统计分析方法从表面上看似乎不属于差异性统计分析方法的范畴,其实,在本质上仍属于此范畴。例如,一致性检验、对称性检验等;最明显的是:非劣效检验、等效性检验和优效性检验这三种特殊的检验方法,它们似乎绝对不属于差异性检验,但在本质上,它们仍然属于差异性检验的范畴。只不过,在这些特殊检验的场合下,前面提及的“差量”不是直接与“0”比较,而是与一个“非 0 的 δ 值”比较而已,这个(在非劣效性检验和优效性检验中,都只涉及一个在临床上有意界的界值 δ , 分别被称为非劣效界值 δ_L 或优效界值 δ_U , δ_L 常取负值、 δ_U 常取正值,正负号仅代表“方向”,在本质上都是一个非 0 的数值)或这些(在等效性检验中,同时涉及两个在临床上有意界的界值,分别称为等效下界值 δ_L 与等效上界值 δ_U)。

可用于单因素两水平设计以下且为一元定量资料的差异性检验方法有:参数检验法包括以正态分布为其理论根据的 Z 检验、以 t 分布为其理论根据的 t 检验(包括用于非劣效检验、等效性检验、优效性检验);非参数检验法包括多种基于不同赋值方法的秩和检验法、经验分布法和精确检验法。

可用于单因素两水平设计以下且为一元定性资料的差异性检验方法有:参数检验法包括以正态分布为其理论根据的 Z 检验(主要用于非劣效检验、等效性检验、优效性检验);半参数检验法包括以 χ^2 分布为其理论根据的各种 χ^2 检验;非参数检验法包括多种基于不同赋值方法的秩和检验法、经验分布法和精确检验法。当然,还应该包含实现一致性检验(也称为 Kappa 检验)、对称性检验等方法。

可用于单因素多水平设计或某种特定的多因素设计且为一元定量资料的差异性检验方法有:参数检验法包括以 F 分布为其理论根据的 F 检验和基于其他诸多不同分布的两两比较方法;非参数检验方法非常少,迄今为止,比较成熟的非参数检验方法基本局限于解决单因素设计一元定量或定性资料的差异性分析之中。

至于多因素某种设计下收集的多元定量资料的差异性检验,目前基本上依靠“Wilk's λ 检验”,常将其转换成 F 检验,故应归属于参数检验法。在单因素两水平设计以下的多元定量资料的差异性检验中,也可采用 Hotelling's T^2 检验,只不过此检验完全可被 Wilk's λ 检验所取代。

21.1.4 相关与回归分析

1. 相关与回归分析之间的关系

相关与回归分析在本质上是两大类统计分析方法,但它们往往被同时用于同一个资料中,因此,还是将它们归属于一大类更有实用价值。

相关分析重在强调变量之间的“相互”关系，而回归分析重在强调变量之间的“依赖”关系。适合进行相关分析的变量，在理论上都应当是随机变量，而且，它们之间的地位是平行的、不存在因果关系。而适合进行回归分析的变量，其用做因变量的变量，在理论上应当是随机变量，用做自变量的变量可以是随机的、也可以是非随机变量。一般来说，自变量是导致受试对象出现某种或某些结果出现的原因，而因变量恰好正是用来度量某种或某些结果的。

在有些实际问题中，人们可以将“相关分析”与“回归分析”决然分开来予以实现；而在另一些实际问题中，人们常习惯于将二者融合在一起，以“回归方程”的形式呈现，但却从“相关”的角度去解读变量之间的关系。例如，在典型相关分析中，对于单个典型变量而言，它是一个“隐变量(不可直接观测的变量)”，是一组性质接近的可测显变量的线性组合(在本质上就是一个回归方程)；而对于级别相同的一对典型变量而言，人们又希望考察它们之间的相关性的大小和方向。

2. 纯相关分析的种类

单纯的相关分析方法有如下几种：简单相关分析(可分为三种，即 Pearson 相关分析、Spearman 秩相关分析和 Kendall's tau-b 秩相关分析)、偏相关分析和复相关分析、典型相关分析、自相关分析(用于时间序列资料)、多项相关分析(用于双向有序属性不同的二维列联表资料，简单的方法可用 Spearman 秩相关分析，引入隐变量的复杂相关分析法被称为多项相关分析)和广义相关分析(基于线性关联矩阵导出的 5 种广义相关系数的定义式)，后面两种在实际应用中很少出现。

3. 纯回归分析的种类

回归分析在统计分析中占有极其重要的地位，因为绝大部分统计分析方法在本质上都属于回归分析的范畴。例如， t 检验、方差分析等方法都是线性模型的特例；即便在看上去与回归分析毫无关系的变量聚类分析中，“类成分(隐变量)”却是由所有显变量标准化后的线性组合；更不用说各种多因素回归分析和其他一些多元统计分析方法了。值得注意的是：通常人们所讲的回归分析，强调的是“单纯回归分析”，即因变量和自变量都是可观测的显变量，而且，采用自变量的线性组合或非线性组合形式来呈现因变量。下面简单介绍单纯回归分析的种类。

(1) 单个计量因变量依赖单个计量自变量的回归分析。

在这种情形下，依据两计量变量之间的变化趋势可分为以下四种具体情形分布处置。第一，反映两计量变量之间关系的散布图呈现出近似“圆饼图”的形状或所有的散点在一条平行于或垂直于横坐标轴的直线上，表明两计量变量之间没有任何直线或曲线关系，这意味着不必进行回归分析；第二，反映两计量变量之间关系的散布图呈现出近似“直线变化趋势(所有的散点在一个不太宽的带内随机地分布着，此带既不平行于也不垂直于横坐标轴)”，这意味着可用一条直线回归方程来描述它们之间的数量关系；第三，反映两计量变量之间关系的散布图呈现出某种初等曲线关系(如指数曲线、对数曲线、幂函数曲线、双曲线、多项式曲线、“S 形”或“倒 S 形”曲线等)，这意味着可用一种被称为曲线直线化的方法来呈现两计量变量之间的曲线关系；第四，反映两计量变量之间关系的散布图呈现出某种非初等曲线关系(如类似于 χ^2 分布或 F 分布概率密度曲线族中的一条曲线或比它们更复杂的曲线)，这意味着两计量变量之间存在着某种难以确定的曲线关系，只能用人们可以想到或可以构造出的某些曲线方程去近似描述它或干脆让计算机去模拟、去逼近。

(2) 单个计量因变量依赖多个计量(“计量”常毫无根据地被删除)自变量的回归分析。

这是实际工作者使用得非常多种的一种回归分析，常被称为“多重线性回归分析”。然而，在实践中，人们又“不假思索”地将“计量自变量”堂而皇之地扩充为任意形式的自变量(通常包括四类，即计量的、二值的、多值有序的、多值名义的)。从基于最小平方原理导出回归系数估计值的过程来看，前述的“扩展”其数理依据是不够充分的，这是回归分析没有很好解决的一大现实难题。

特别说明：以下所有在自变量之前冠以“计量”二字的场合下，实际使用时，目前都假定可以将其删除，即允许任何形式的自变量。

当单个计量因变量的取值可以充满(0,1)区间(如每个个体身上可以观测到某种细胞的阳性率)时，就需要基于一定假设成立的条件下建立一元多重回归方程，其特例就是 Probit 多重回归分析。

多个计量自变量并非总是以“线性”形式与计量因变量建立联系，在很多场合下，却以“非线性”形式出现。此时，非线性的形式是很难确定的。通常，只能是用各种猜想的形式或结构去套用现实的资料，在一定精度和标准的前提下，找到相对合适的表达形式并给出回归系数的估计结果。

(3) 多个计量因变量依赖单个计量自变量的回归分析。

这是所谓的多元一重线性回归分析问题，此问题是下面情形的特例。

(4) 多个计量因变量依赖多个计量自变量的回归分析。

这是所谓的多元多重线性回归分析问题，就是在一个实际问题中，同时存在多个计量因变量，它们都依赖同一组计量自变量的变化而变化，将它们视为一个整体，同时建立多个计量因变量依赖同一组多个计量自变量变化的回归方程。其回归系数的参数估计值的求法常有一般最小平方法和偏最小平方方法两种。

路径分析实际上就是一种特殊的多元多重线性回归分析，它与前面所述及的多元多重线性回归分析相比，其不同点是在作为因变量的变量中有两种情形：其一，是中间变量；其二，是“纯因变量”。纯因变量只能作为某些自变量的因变量，而那些中间变量在某些方程中只能作为自变量，而在另一些方程中也可作为因变量。

(5) 单个计数因变量依赖多个计量自变量的回归分析。

这是所谓的一元多重 Poisson 回归分析问题，一元一重 Poisson 回归分析是其特例。此类回归分析最大的特点是因变量为基数，例如，在某个实际问题中，因变量为不同条件下(即自变量不同水平组合)某事物或现象出现的次数，若随着条件的改变，因变量的全部取值近似服从 Poisson 分布时，建立因变量随自变量变化而变化的依赖关系，可选择此种回归分析。当发现各观测点上因变量取值的离散度很大(称为过离散)时，就需要对回归模型进行修改，通常改用负二项回归分析。

(6) 单个生存时间变量依赖多个计量自变量的回归分析。

这是所谓的生存资料回归分析问题。生存资料有两个明显特点：第一，生存资料中存在不完全的数据(也称为删失数据或截尾数据)；第二，生存时间通常偏离正态分布很远。

有鉴于此，普通的统计分析方法已不再适合用来分析生存资料。取而代之的是生存分析。

生存分析方法通常可以分为四类：第一类，可以分析单因素生存资料的非参数统计分析方法，常用的有乘积极限法(也称为 Kaplan-Meier 法)或寿命表法；第二类，Cox 比例风险模型分析法(半参数模型分析法之一)；第三类，Cox 非比例风险模型分析法或称为含时变协变量的 Cox 风险模型分析法(半参数模型分析法之二)；第四类，参数模型分析法(当生存时间所服从的某种分布能被准确地鉴定出来且有相应的建模方法，如对数正态分布模型、威布尔分布模型等)。

值得一提的是：在对生存资料建模时，并不是直接以生存时间 t 为因变量，而是以 t 的某种函数为因变量，通常以危险率函数 $h(t)$ 为“因变量”，建立此种因变量关于全部自变量的依赖关系。

(7) 单个时间序列数据的回归分析。

当人们关心的事物或现象随着时间的推移，在整体上在特定的时间点上出现的数量大小是随机的，但的确又存在一定的规律性时，就可称其为时间序列数据。例如，每年天空中某指定区域内出现的太阳黑子数目；全世界每年某急性传染病的发病率等。若仅呈现某数量指标对应的一串时间序列数据时，属于简单的时间序列数据；若在呈现时间序列数据的同时，还收集了那些可能与时间序列数据有关的其他因素的数据，就属于多因素时间序列问题。

研究一串时间序列数据前后之间相依关系的方法就称为时间序列分析，在本质上，就是一种特殊的回归分析。根据时间序列数据之间关系的复杂程度不同，也可分为线性时间序列分析与非线性时间序列分析。

(8)单个定性因变量依赖一个或多个计量自变量的回归分析。

定性因变量有三种表现：第一，二值因变量；第二，多值有序因变量；第三，多值名义自变量。这三种场合下对应的回归分析都称为多重 logistic 回归分析，其区别是回归方程的复杂程度不同。分别称为因变量为二值变量的多重 logistic 回归分析(对于某种特定的结局而言，只有一个普通的多重 logistic 回归方程)、因变量为多值有序的多重累计 logistic 回归分析(对于某种特定的结局而言，有 $(k - 1)$ 个普通的多重 logistic 回归方程， k 为多值有序变量的水平数，它们的截距项不同，但相同自变量前的回归系数是相同的)、因变量为多值名义的多重扩展 logistic 回归分析(对于某种特定的结局而言，有 $(k - 1)$ 个普通的多重 logistic 回归方程， k 为多值名义变量的水平数，它们的截距项不同且相同自变量前的回归系数也不相同，而且，logistic 曲线回归方程的分母上的指数函数的项数也不同)。

(9)以隐变量为因变量的回归分析。

绝大多数多元统计分析方法所对应的数据结构为“单组设计多元定量资料”，即资料仅涉及“单因素—水平多个结果变量”。在实际处理资料时，常将“对个因变量”视为原因变量或自变量，而将那些“躲藏”在它们背后间接起作用的变量(统计学上称其为隐变量)视为因变量，构造出“回归方程”，例如，主成分分析中的主成分表达式、变量聚类分析中的类成分表达式、探索性因子分析中的因子得分模型表达式、典型相关分析中的典型变量表达式、判别分析中的判别表达式等。

21.1.5 广义综合评价

1. 经典的综合评价

(1)无层次结构的经典综合评价。

某研究者收集了东南大学附属中大医院 1990—1999 年住院医疗质量资料，见表 21-1，试对各年度的 9 个医疗质量指标进行综合评价。

表 21-1 东南大学附属中大医院 1990—1999 年住院医疗质量指标

年度	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1990	99.3	99.8	99.9	99.7	72.0	94.1	71.2	1.8	20.7
1991	99.1	99.6	99.7	99.4	73.8	93.9	81.5	2.1	21.2
1992	99.2	99.6	99.7	99.5	70.5	94.3	86.0	1.7	20.0
1993	99.5	99.7	99.9	99.9	71.9	94.3	85.2	2.1	19.7
1994	99.3	99.5	99.7	99.6	73.2	94.6	86.1	2.1	19.1
1995	99.4	99.5	99.8	99.8	73.2	94.2	89.7	2.1	17.9
1996	99.7	99.7	99.9	99.8	71.0	88.0	90.6	1.8	16.5
1997	99.7	99.7	99.8	99.7	72.2	94.9	92.9	1.9	16.0
1998	99.8	99.8	99.8	100.0	66.1	94.6	94.9	2.1	15.2
1999	99.6	99.8	99.9	99.9	67.2	94.8	94.0	2.0	14.9
w	0.102	0.111	0.131	0.094	0.091	0.079	0.114	0.084	0.194

注： $x_1 \sim x_8$ 分别表示门诊与出院符合率(%)、入院与出院符合率(%)、术前后诊断符合率(%)、临床病理诊断符合率(%)、治愈率(%)、好转率(%)、抢救成功率(%)、病死率(%)， x_9 表示平均治愈天数。 w 表示各指标权重。

本例中观测了东南大学附属中大医院在 10 年里 9 项与医院医疗质量有关指标的取值，这些指

标所测自的受试对象(即该医院)来自一个层次,仅在“时间”因素进行了“重复测量”,本质上是性质相同或接近的 9 个时间序列数据的“平行展示”。若统计分析的目的是希望基于 9 项定量指标来给该医院在 10 年里的表现好坏进行综合评价,可以采用多种多元统计分析方法(如主成分分析、探索性因子分析等)来间接实现,但在样本含量较少小,人们更习惯于采用简单的非参数统计分析方法,如秩和比法、熵值法、Topsis 法等,它们被统称为无层次结构的经典综合评价方法。

(2)有层次结构的经典综合评价。

某市领导拟对该市 6 所综合医院的工作质量进行评估,主要以医疗工作、护理工作和膳食供应三个方面来进行评价,这三个方面均通过相关详细指标来反映。医疗工作通过医疗制度、医疗质量和病床使用情况来反映,护理工作通过护理制度来反映,膳食供应通过膳食质量来反映。另外,医疗质量可细分为重症收容和一般治疗疗效两方面。有关指标数据见表 21-2 和图 21-1。

表 21-2 某市 6 所医院工作质量评价情况

医院	病床使用率	治疗有效率	重症收容率	医疗制度 执行优良率	护理制度 执行优良率	膳食供应 优良率
A	95.0	88.1	15.4	74.7	54.7	41.3
B	92.0	91.2	8.3	53.4	20.7	41.4
C	94.8	90.0	7.9	61.9	26.1	22.8
D	95.6	94.0	3.1	50.0	20.0	20.0
E	89.1	93.6	9.5	61.9	27.4	34.0
F	77.4	92.2	3.7	67.1	35.5	30.3

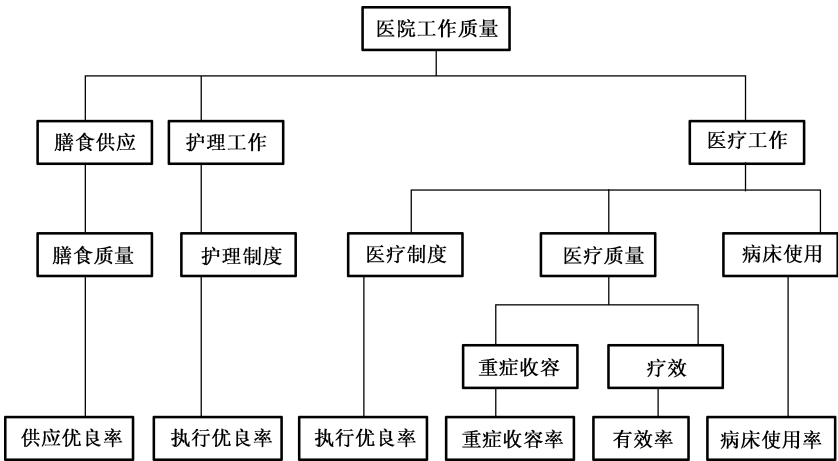


图 21-1 具有三个层次的医院医疗质量指标体系

由图 21-1 可知:供应优良率、护理制度执行优良率、医疗制度执行优良率、病床使用率这 4 个指标仅具有两级结构,而重症收容率和治疗有效率这两个指标却具有三级结构。

对于这种指标具有多层次结构的管理类科研资料,若希望得到综合性的评价指标,并据此对全部 6 所医院进行优劣排序或分档,就需要采用层次分析法这种综合评价方法。

2. 非经典的综合评价

(1)差异性分析。

当研究者收集到多项同类研究,其中每项研究都属于一个差异性分析问题,就可运用 meta 分析将这多项同类研究合并在一起得出一个概括性的差异性分析的结论。而有时,多项同类研究是

基于一个定量指标数值由小到大取不同数值而产生的,此时,不适合采用 meta 分析将多项研究合并在一起(因为这些研究项目之间可能存在本质区别),而需要采用 ROC 曲线分析法,用动态的眼光寻找到差异非常明显或区分度很大(同时参考灵敏度于特异度量项诊断指标)定量诊断指标的“分割点(cut-off point)”。

(2)直接或间接实现样品聚类分析。

在表 21-1 中,若将 10 个年份视为 10 个个体,9 项指标仍视为从每个个体身上观测的 9 项定量指标,人们可以采用“无序样品聚类分析”将 10 个个体分成若干类;也可以运用主成分分析和探索性因子分析等多元统计分析方法间接实现样品聚类分析;若依据各变量的取值求出任何两个个体之间的“距离”并构造出“相似度矩阵”,还可采用多维尺度分析,将全部个体之间的相对位置关系在二维平面上呈现出来。

(3)直接或间接实现变量聚类分析。

采用变量聚类分析可以直接实现变量聚类;而采用典型相关分析、主成分分析和探索性因子分析可间接实现变量聚类。

(4)同时研究变量与样品之间可能存在的某些关系。

在表 21-1 中,若人们关系的问题是:10 个个体与 9 个指标之间是什么关系呢?这个问题的更具体的问法是:哪些个体在哪些指标上的取值规律相同或接近?可以采用对应分析法来比较好地回答前述提出的问题。

21.2 合理选用统计分析方法的关键技术

21.2.1 合理选用统计分析方法的四要素

当有理由认为所给定的科研数据是值得分析时,接下去的一个关键性问题是如何才能合理选用统计分析方法。要想做到这一点,必须明白:合理选用统计分析方法有四个要素,它们是“问题与数据结构”、“统计分析目的”、“由统计分析方法定义的数据子集”和“数据子集与拟选用的统计分析方法所要求的前提条件是否吻合”。

(1)问题与数据结构。

它们是统计分析的“对象”,在选用统计分析方法之前,必须对分析“对象”有一个全面而深入的了解,否则,就将是“盲人打靶”、无的放矢。

(2)统计分析目的。

对于给定的科研资料,可能存在多个不同的统计分析目的。若分析目的不明确,就无法或无依据去选择某种具体的统计分析方法。稍不留神,就可能南辕北辙、背道而驰。

(3)由统计分析目的定义的数据子集。

若科研数据以数据库的形式呈现出来,一般来说,其行数就是样本含量(除重复测量设计资料外),而其列数接近拟参与统计分析的变量数(一般指“原因变量”和“结果变量”;排除不参与统计分析的“编号”和“家庭住址”或“联系方式”等一般变量)。与每一个具体的统计分析目的对应的数据,通常是全分析集中的一部分。例如,一个科研资料包含正常人和某病患者相同的一组变量的取值,当统计分析目的为“研究正常人各指标之间的内在关系”时,就从全分析集中删除某病患者对应的那些行上的数据,就构成了与此分析目的对应的一个统计分析的数据子集;当统计分析目的为“研究正常人与某病患者在 5 项血脂指标上的差异是否具有统计学意义”时,就可以从全分析集中删除其他变量所对应的那些列。

(4)数据子集与拟选用的统计分析方法所要求的前提条件是否吻合。

数据子集中的观测数目是否符合拟选用的统计分析方法所要求的“样本含量”？数据子集中的某些变量的取值是否满足拟选用的某些统计分析方法所要求的前提条件(例如，若是某种设计下收集的定量资料，选用参数检验法分析资料时，要求资料应满足独立性、正态性和方差齐性；若是列联表资料，一般 χ^2 检验要求列联表资料中小于 5 的理论频数的个数不应超过此列联表中总格子数的 1/5)。

把握好以上“四个要素”，轻松实现合理选择统计分析的目的就是一件水到渠成的事了。

21.2.2 合理选用统计分析方法的实例演示

下面沿用第 20 章中的一个例子及其表格中的数据，为了避免读者前后翻书所带来的不必要麻烦，将那个例子和数据复制在下面，并重新编号。

1. 问题与数据结构

有人对 103 例冠心病患者($G=1$)和 100 例正常对照者($G=2$)进行了多项指标的观测，资料见表 21-3。试依据上述的“合理选用统计分析方法的四要素”，举例说明之。

表 21-3 冠心病人与正常人多项指标的观测结果

编号	组别	性别	年龄	高血压史	吸烟史	胆固醇	甘油三脂	低密度脂蛋白	高密度脂蛋白	脂蛋白 α	载脂蛋白 As	载脂蛋白 B	基因型 xbaI	基因型 EcoRI	服药情况
N	G	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
1	1	男	60	无	无	223	205	122	30	106	0.92	0.74	-/-	-/-	未服
2	1	女	46	无	无	166	51	84	57	56	1.14	0.54	-/-	+/-	β 阻滞剂
3	1	男	55	有	无	273	155	197	34	58	1.01	0.93	+/-	+/+	β 阻滞剂
4	1	男	30	无	有	193	99	70	37	85	0.94	0.70	-/-	+/+	降血脂药
...
102	1	女	66	无	无	194	77	107	54	43	0.98	0.61	-/-	+/+	未服
103	1	男	76	有	有	195	101	135	42	302	0.91	0.65	-/-	+/+	β 阻滞剂
104	2	女	39	无	无	223	48	157	55	32	1.06	0.71	-/-	+/+	未服
105	2	男	40	无	无	152	91	100	46	62	1.00	0.55	-/-	-/-	未服
106	2	男	59	无	有	168	84	106	33	205	0.67	0.69	-/-	+/+	未服
...
203	2	男	69	有	无	224	110	58	49	132	1.10	0.96	-/-	+/+	未服

2. 统计分析目的

根据研究者或分析者的需要，可以提出许许多多不同的统计分析目的。现假定有人提出了下面四个不同的分析目的。

- (1) 第一个统计分析目的：分别按“冠心病患者”与“正常人”了解各个变量的频数分布规律。
- (2) 第二个统计分析目的：仅考察“冠心病患者”与“正常人”在 X_5 这一项血脂指标之间的差别是否具有统计学意义，但希望尽可能消除年龄和性别所造成的不利影响。
- (3) 第三个统计分析目的：同时考察“冠心病患者”与“正常人”在七项 ($X_5 \sim X_{11}$) 血脂指标之间的差别是否具有统计学意义。
- (4) 第四个统计分析目的：重点希望揭示冠心病患者七项 ($X_5 \sim X_{11}$) 血脂指标之间的相互关系和依赖关系。

3. 由统计分析目的定义的数据子集

下面依据前面提出的四个不同的统计分析目的，分别给出由它们定义的数据子集。

(1)与上述第一个分析目的对应的数据子集。删除全分析集中的前两列和最后一列后所得到的那些数据组成的数据子集,但是,每次提出数据信息时,要么选用前 103 行与“冠心病患者”对应的所有各列数据中的某一列;要么选用后 100 行与“正常人”对应的所有各列数据中的某一列。

(2)与上述第二个分析目的对应的数据子集。在全分析集中,仅保留表中“ G 、 X_1 、 X_2 、 X_3 ”四列数据后所得到的那些数据组成的数据子集。

(3)与上述第三个分析目的对应的数据子集。在全分析集中,仅保留表中“ G 、 $X_3 \sim X_{11}$ ”八列数据后所得到的那些数据组成的数据子集。

(4)与上述第四个分析目的对应的数据子集。在全分析集中,仅保留表中前 103 行且“ $X_3 \sim X_{11}$ ”七列数据后所得到的那些数据组成的数据子集。

4. 数据子集与拟选用的统计分析方法所要求的前提条件是否吻合

有了统计分析目的和与之对应的数据子集,再结合统计分析方法的分类知识,就可以初步确定拟选用的大致统计分析方法所在的“种类”,检查或判断数据是否满足备选统计分析方法中某些方法所要求的前提条件。分别根据资料是否满足特定的前提条件,最终确定应选择什么具体的统计分析方法。

(1)与上述第一个分析目的对应的统计分析方法的选择。这是探索性统计分析的目的,只需注意选择定性还是定量的统计表达与描述方法来呈现资料即可。

(2)与上述第二个分析目的对应的统计分析方法的选择。应检查年龄作为自变量、 X_3 作为因变量所得到的回归直线在由性别与受试者种类所形成的 4 个组之间是否分别平行。若平行,则可选用具有一个协变量(即年龄)的两因素析因设计(注:非标准的析因设计)定量资料一元协方差分析处理资料;否则,可能需要考虑采用非线性混合效应模型处理了。

(3)与上述第三个分析目的对应的统计分析方法的选择。若资料在两种受试对象之间满足方差协方差矩阵齐性要求,可选择单因素两水平设计定量资料七元方差分析处理;否则,就应设法寻找到合适的变量变换方法使其满足前提条件,再对变换后的数据采用前述的多元方差分析处理。

(4)与上述第四个分析目的对应的统计分析方法的选择。此时,数据结构被称为单组设计七元定量资料。可以选用的多元统计分析方法很多,它们一般对资料没有特殊的要求,但最好全是定量资料(在实际使用中,也有人盲目将定性资料混杂在其中)。可选择的多元统计分析方法如下。

- ①在不提供任何附加信息的前提下,可选择主成分分析、探索性因子分析、变量聚类分析、无序样品聚类分析、对应分析。
- ②在提供少量附加信息的前提下,可选择单组设计定量资料多元方差分析、典型相关分析、路径分析、证实性因子分析和结构方程模型分析。
- ③在提供较多附加信息(相似度矩阵或不相似度矩阵)的前提下,可选择多维尺度分析。

(胡良平 胡纯严 高 辉 刘一松 陈 娟 贾玉凤)

第 22 章 临床科研统计设计与数据分析 常见差错辨析与释疑要览

人们在进行临床科研设计和数据分析中,经常会犯各种各样的错误,例如,忽视伦理道德问题、选择了不合适的受试对象、盲目地决定采用多少样本含量、未制定合适的选择受试对象的质量标准就盲目选择受试对象、未找准找全试验因素和重要非试验因素、未对重要非试验因素进行严格的质量控制、未选准反映试验效应的指标、对指标未进行合理的分类、未对指标的测定方法和测定时间进行标准化、未选用合适的方法对受试对象进行随机抽样或随机分组、未设置出合理的对照组、忽视重复原则(未做到有根据地给出样本含量)、各对比组之间通常不满足均衡性要求、缺乏试验设计类型知识(常以分几组取代相应的某种设计类型)、常错误理解比较类型(例如,未弄清“非劣效检验”的含义而盲目套用此方法;搞不清楚等效性检验与双侧差异性检验之间的区别;混淆了单侧差异性与优效性检验之间的概念)。诸如此类的错误不胜枚举,本章将从一些公开发表的临床学术论文中找出一些典型案例,对其进行全面的辨析与释疑,以便提高读者识别临床学术论文中常见统计学错误的能力。

22.1 康复医学研究论文统计学差错辨析与释疑

22.1.1 基于单因素 4 水平设计的临床试验研究

原文题目: Glazer 盆底表面肌电评估方案在脊髓损伤患者中的应用。

1. 原文主要内容摘要及对存在问题的剖析

在与“伦理学”有关的内容方面,原文中未提及此项临床试验研究是否得到了有资质的伦理委员会的审批,也未提及是否与患者签订了知情同意书。

在与临床试验“三要素”有关的内容方面:(1)在“受试对象”方面,①在确定受试对象种类方面,原文中为评估不同损伤程度的脊髓损伤(SCI)患者肛门括约肌表面肌电的变化,因此选取 2011 年 1—9 月在中国康复研究中心住院并申请水疗的 SCI 患者 110 例作为研究对象,但是并未说明选择受试对象的理由;②在制定受试对象“质量标准”方面,原文中明确写出了“排除标准”,但未见“纳入标准”、“中止标准”、“终止标准”和“剔除标准”;③在确定受试对象“数量标准”方面,没有给出确定样本含量的依据。(2)在与“影响因素”有关的内容方面,①原文中未明确指出试验因素是什么;②未找准找全对主要结果有影响的重要非试验因素。(3)在与“试验效应”有关的内容方面,①在观测指标的分类方面,叙述含糊,未明确分出诊断指标、疗效评价指标和安全性指标;②在观测指标的主次方面,为反映试验效应,原文中罗列了很多观测指标,但未指出主次指标分别是什么;③在指标的测定方法、测定次数和测定时间等方面,虽有所提及,但交代得都不够十分清晰。

在与临床试验“四原则”有关的内容方面:(1)在“随机原则”方面,直接对 110 例患者根据 A-SIA 残损分级进行分组,即 A 级、B 级、C 级和 D 级的患者分别分为 A 组、B 组、C 组和 D 组 4 个组。这种确定式分组,违背了统计学上要求随机抽取受试对象的基本要求。由于本例中的“残损分

级”是一个与受试对象直接有关的“属性因素”，无法对全部被选定的受试对象进行随机化分组，唯一能体现随机化的做法就只能从总体内 A 级、B 级、C 级和 D 级的患者中分别随机抽取指定样本含量的患者，但临床医生几乎都无法这样做，只是选用到自己医院来就诊的数量十分有限的那些患者。由此可知，相当多的临床研究是不符合随机化原则要求的。(2)在“对照原则”方面，原文中未明确提及所采用的“对照形式”，但不难看出，采取的是残损 A 级、B 级、C 级和 D 级之间的“相互对照”。就本例的研究目的而言，采用相互对照是合理的。(3)在“重复原则”方面，特指在相同试验条件(即每个小组)下应进行足够多次数的独立重复试验，本例毫无根据地给出总样本含量为 110 例，再按残损级别分为 4 级，各级例数分别为 45、19、20、26 例。虽然各组例数不算很少，但没有任何证据表明是严格遵守“重复原则”的。(4)在“均衡原则”方面，强调各对比组之间在一切非试验因素方面都几乎一致。本例按残损级别分为 4 级后，没有对其他任何非试验因素采取任何措施，故均衡性时很差的。

在与临床试验“设计类型”有关的内容方面：原文中只是将患者按残损级别分为 4 级，对应于 A、B、C、D 四组，未明确交代此项临床试验研究采用的是什么设计类型。就本例而言，实际上对应的设计类型为“单因素 4 水平设计”。

在与临床试验“比较类型”有关的内容方面：本例属于临床试验研究方法的疗效评价问题，而不是新药或新医疗器械的疗效与安全性的评价研究，通常都采用差异性检验，而不需要采用其他的比较类型(非劣效检验、等效性检验和优效项检验)。

在与“质量控制”有关的内容方面，究竟在此项临床试验研究过程中涉及哪些必须加以控制的非试验因素，以及如何有效地去控制它们，在原文中被遗漏掉了。临床试验结果受到多种非试验因素的影响(尤其是来自研究者和受试者心理因素、环境与条件、指标的测定方法与技术水平)，质量控制做得好坏，直接关系到临床试验研究所得到的资料是否准确可靠、是否值得进行统计分析。

在与“资料表达”有关的内容方面，原文中提供了一些以简化形式呈现的临床资料，其中一张统计表复制成如表 22-1 所示。

表 22-1 4 组间 sEMG 检测值两两比较(μV, $\bar{x} \pm s$)

组别	列数	Pro-Mean	Flick-Max	Tonic-Max	Tonic-Mean	EduL-Max	EduL-Mea
A 组	45	5.37 ± 6.58	9.45 ± 9.77	9.13 ± 10.04	4.80 ± 5.49	9.26 ± 10.25	4.06 ± 4.47
B 组	19	6.70 ± 6.69	12.58 ± 16.00	11.76 ± 14.52	6.95 ± 8.09	12.53 ± 17.30	6.43 ± 8.20
C 组	20	5.34 ± 4.48	37.16 ± 26.01 ^b	30.49 ± 19.56 ^b	14.62 ± 8.65 ^b	23.15 ± 15.77 ^b	1.53 ± 6.71 ^b
D 组	26	7.34 ± 5.92 ^a	62.67 ± 36.70 ^{bc}	56.83 ± 34.39 ^{bd}	29.18 ± 17.65 ^{bd}	43.82 ± 21.90 ^{bd}	22.38 ± 11.54 ^{bd}
组别	列数	Edu2-Max	Edu2-Mean	Edu3-Max	Edu3-Mean	Post-Mean	
A 组	45	7.64 ± 8.42	3.73 ± 4.13	6.96 ± 7.81	3.80 ± 4.48	2.72 ± 3.34	
B 组	19	12.72 ± 15.80	6.10 ± 8.04	9.62 ± 13.60	5.97 ± 7.94	5.02 ± 5.51	
C 组	20	25.17 ± 15.08 ^b	10.35 ± 6.92 ^b	18.42 ± 12.35 ^b	9.89 ± 7.04 ^b	3.98 ± 2.61 ^b	
D 组	26	42.57 ± 27.62 ^{bc}	20.85 ± 13.05 ^{bd}	36.07 ± 23.37 ^{bd}	20.24 ± 13.11 ^{bd}	5.62 ± 4.43 ^b	

注：与 A 组比较，^a代表 $P < 0.05$ ；^b代表 $P < 0.01$ ；与 C 组比较，^c代表 $P < 0.05$ ，^d代表 $P < 0.01$ 。

由此表中的数据可看出，有不少组中定量指标的标准差大于其算术平均值，说明这些组中的原始数据不呈正态分布，改用“ $M(Q_1 \sim Q_3)$ ”的简化形式呈现数据更合适。

在与“统计分析”有关的内容方面，原文有如下的一段文字：采用 SPSS 13.0 版软件对患者的各项检测指标进行统计分析。因为 4 个组的样本例数相差较大，故采用多个独立样本非参数检验方法(Kruskal Wallis Test)分析 4 个组之间的各检测阶段 sEMG 的最大值及平均值。采用两独立样本非参数检验方法(Mann-Whitney Test)进行各组之间的两两比较。采用配对两样本秩和检验方法分析 110 例 SCI 患者的前基线测定阶段和后基线测定阶段检测指标的差异。

这段文字反映出了不少问题：首先，选择非参数法分析原文资料的理由并不充分；所采用的两两比较方法未控制犯第 I 类错误发生的概率；在每一组内部采用自身配对设计也同样会增大犯阳性错误的概率。

2. 对原文中存在的主要问题的辨析与释疑

原文中所反映出来的问题表面上是“论文写作中的问题”，其实，是绝大多数临床试验研究工作的“通病”。这个通病可概述如下。

缺乏正确的“统计思想”的指导，不会通过事物或现象的“表现型”去找出其“原型和标准型”，不知道临床试验研究的“核心内容”是什么，用“回顾性总结临床病例”取代“前瞻性临床试验研究”。其后果是：大多数临床试验研究只有“苦劳”而没有“功劳”，临床数据的准确性不高、结论的可信度不高、对未临床研究和应用实践的指导价值不大。

本书中从第 1 章到第 18 章已介绍了很多统计思想、三型理论、统计理论和方法，此处不再赘述。这里只想强调一点，那就是任何临床试验研究项目，首先必须制订出科学完善经济可靠的临床试验设计方案、编制出内容全面且言简意赅的临床病例报告表；其次，应自始至终坚持在“临床试验设计方案”的指导下，实时严格地进行“质量控制（包括采取盲法消除或降低来自研究者和受试者心理因素对结果的影响、尽可能消除或降低环境和条件对结果的影响、在多中心临床试验研究中应设法降低中心效应）”；最后，应进行科学的数据管理和统计分析。

临床试验设计的核心内容必须得到充分体现，应做到尽善尽美。它们是：伦理道德问题、临床试验设计三要素与四原则、设计类型与比较类型、质量控制与数据管理。

22.1.2 基于具有一个重复测量因素的两因素设计的临床试验研究

原文题目：不同运动方式对乳腺癌术后患者上肢功能及生活质量的影响。

1. 原文主要内容摘要及对存在问题的剖析

这篇原文显示出：在临床试验设计的核心内容上与上一篇原文大同小异，有所改善的地方有以下两点。（1）制定了比较全面的受试者质量标准。具体的摘录是，患者纳入标准如下：①长期居住芜湖地区；②年龄为 18 ~ 65 岁；③病理检查证实为乳腺癌，并进行改良根治手术治疗；④患者对本研究知情同意。患者排除标准如下：①预后较差的晚期乳腺癌或非原发性乳腺癌患者；②合并有严重的心脑血管疾病或精神病史。

（2）采取了比较正规的方法对受试对象进行了随机化分组。具体的摘录如下：采用随机数字表法将 185 例乳腺癌患者分为太极拳组、秧歌舞组及对照组，即采用了完全随机化方法对全部纳入患者进行了分组。

原文用表格形式呈现了基线资料和以简化形式表达的临床试验研究结果，现将其中一张表格复制成如表 22-2 所示。

原文处理多张类似表 22-2 的资料的统计分析方法摘录如下：本研究所得计量数据以 $(\bar{x} \pm s)$ 表示，采用 SPSS 12.0 版统计学软件包进行数据分析，计量资料比较采用 t 检验，计数资料比较采用 χ^2 检验， $P < 0.05$ 表示差异具有统计学意义。

2. 对原文中存在的主要问题的辨析与释疑

这篇原文在临床试验设计和数据统计分析等方面所犯的错误与前面介绍的第一篇原文大同小异，此处不再一一赘述。下面仅就数据的呈现和统计分析方法的选择方面存在的问题进行辨析与释疑。

（1）表 22-2 编制得不符合要求。给读者的感觉是该临床试验研究仅考察一个试验因素，它的名

字称为“组别”。其实，“组别”是由两个试验因素复合而成的。一个名叫“锻炼方式”的试验分组因素、另一个名叫“观测时间”的重复测量因素。由于列表的错误，很容易使人误判此临床试验资料所取自的设计类型为“单因素 9 水平设计”，而正确的试验设计类型为“具有一个重复测量因素的两因素设计”。这一误判，导致了原文作者错误地选用了统计分析方法。正确的列表格式如表 22-3 所示。

表 22-2 术后不同时间点各组患者上肢功能评分情况比较(分, $\bar{x} \pm s$)

组别	例数	疼痛	日常行动功能	关节活动	肌力	总分
太极拳组						
术后 10d 时	63	9.38 ± 2.67	7.14 ± 1.26	14.75 ± 3.24	12.29 ± 2.72	43.50 ± 6.14
术后 1 月时	63	13.32 ± 3.29	10.65 ± 3.15 ^{ab}	25.58 ± 4.16 ^{ab}	18.13 ± 3.96	67.17 ± 11.51 ^a
术后 3 月时	63	14.54 ± 3.23 ^a	18.78 ± 3.82 ^{ab}	35.84 ± 5.04 ^{ab}	22.58 ± 6.18	91.38 ± 12.28 ^a
秧歌舞组						
术后 10d 时	51	9.30 ± 2.49	7.05 ± 1.59	14.81 ± 3.39	12.35 ± 2.56	43.49 ± 5.93
术后 1 月时	51	12.89 ± 3.37	9.59 ± 3.11	24.02 ± 4.47	17.49 ± 3.28	64.49 ± 12.03
术后 3 月时	51	13.77 ± 3.52	17.49 ± 3.66 ^a	33.55 ± 5.09	22.15 ± 5.24	87.89 ± 12.66 ^a
对照组						
术后 10d 时	71	9.24 ± 2.57	7.06 ± 1.46	14.91 ± 3.73	12.43 ± 2.31	43.53 ± 6.85
术后 1 月时	71	12.42 ± 3.06	9.24 ± 3.35	23.84 ± 4.26	17.58 ± 3.51	63.27 ± 10.19
术后 3 月时	71	13.37 ± 3.29	16.17 ± 4.04	32.26 ± 5.72	21.31 ± 5.77	83.11 ± 13.75

注:与出院时比较,^a代表 $P < 0.05$;与秧歌舞组相同时间点比较,^b代表 $P < 0.05$ 。

表 22-3 术后不同时间点各组患者上肢功能评分情况比较(分, $\bar{x} \pm s$)

锻炼方式	疼痛评分			…	总分		
	术后 10d 时	术后 1 月时	术后 3 月时		术后 10d 时	术后 1 月时	术后 3 月时
太极拳组	9.38 ± 2.67	13.32 ± 3.29	14.54 ± 3.23		43.50 ± 6.14	67.17 ± 11.51	91.38 ± 12.28 ^a
秧歌舞组	9.30 ± 2.49	12.89 ± 3.37	13.77 ± 3.52		43.49 ± 5.93	64.49 ± 12.03	87.89 ± 12.66
对照组	9.24 ± 2.57	12.42 ± 3.06	13.37 ± 3.29		43.53 ± 6.85	63.27 ± 10.19	83.11 ± 13.75

注:优点是设计类型一目了然;缺点是表格变得很宽。

(2)误判了试验设计类型而导致错误地选择了统计分析方法。原文作者选用 t 检验处理其原文中类似表 22-2 中的计量资料是错误的。因为在进行算术平均值之间差异性分析时,仅适合满足参数检验前提条件时三种设计类型下的一元定量资料,即单组设计、配对设计和单因素两水平设计(常被简称为成组设计)。本例对每项定量指标而言,宜选用具有一个重复测量因素的两因素设计定量资料一元方差分析(可借助 SAS 软件包中的混合效应模型过程 MIXED 来实现)。

值得一提的是:原文的统计分析部分还有这样一句话,即“计数资料比较采用 χ^2 检验”。但仔细阅读原文,发现原文作者所指的“计数资料”是接受三种不同锻炼方式的患者在三个年龄段上的频数构成情况、婚姻状况(未婚与已婚)频数构成情况、三种疾病类型频数构成情况和三种疾病分期频数构成情况。其实,这些资料本质上都属于“定性资料”。这是大多数毕业于医学院校的人们学习医学统计学或卫生统计学时被“误导”的结果,换句话说,早年的医学或卫生统计学教科书上就一直把“定性资料”错误地定义成“计数资料”。

22.2 非康复医学研究论文统计学差错辨析与释疑

22.2.1 基于多重线性回归分析的临床试验研究

原文题目:多重线性回归分析自身因素对全膝关节置换术后疗效的影响。

1. 原文主要内容摘要及对存在问题的剖析

如 22.1 节所示，这篇原文在临床试验设计方面同样存在不少问题，此处不再赘述。下面仅复制该文中的一张统计表，见表 22-4；同时，将原文中关于“统计学分析”部分内容复制如下：

数值变量资料以 $\bar{x} \pm s$ 表示，两组间均数比较采用 t 检验，多组均数间比较采用方差分析(ANOVA)，均数间两两比较采用 SNK 法，所有资料检验均为双侧($\alpha = 0.05$)；方差齐性检验($\alpha = 0.10$)；采用多重线性回归分析评估自身相关因素对术后疗效(术后 HSS 评分)的影响。数据采用 SAS 8.0 统计软件进行分析(注：HSS 代表特种外科医院)。

表 22-4 自身因素对术后 HSS 评分影响的多重线性回归分析

变量	参数估计	标准误	t	P	标准偏回归系数
截距	-2.29483	1.45986	-1.58	0.1583	0
年龄	-0.03717	0.00278	-1.13	0.0967	-0.35491
性别	0.03339	0.02923	0.92	0.0802	0.34712
BMI	0.00507	0.02923	0.92	0.0802	0.34712
病因	0.06348	0.03245	2.09	0.0379	0.45655
术前状况(分)	0.07253	0.02423	2.32	0.0254	0.50461

由表 22-4 倒数第 2 列的“ P 值”不难看出：只有最后两个自变量对应的 P 值小于 0.05，说明所求的多重线性回归方程的质量是不高的，据此作出任何统计和专业结论是不可信的。

2. 对原文中存在的主要问题的辨析与释疑

原文除了在临床试验设计方面存在很多前文提及的“通病”之外，在进行多重线性回归的全过程中存在很多致命的错误，现着重论述如下。

(1) 对定量结果变量 HSS 可能有影响的因素很多，原文作者仅选取了很少的几个，离“找准找全”影响因素的要求相差甚远。

(2) 认为地将原先的定量变量“年龄”、“BMI”和“术前评分”都分别划分成三挡，是否科学合理未做解说。

(3) 将三个多值有序的自变量(年龄、BMI、术前评分)简单地赋值 1、2、3，其科学合理性未做证实。

(4) 当所研究的问题中存在多个自变量时，创建多重线性回归方程时未对自变量进行筛选，导致最终的结果中包含了很多无统计学意义的项(包括截距项)。

(5) 未对自变量之间是否存在多重共线性进行诊断、也未对观测点进行异常点诊断，无法保证所创建的多重线性回归方程的质量。

22.2.2 基于生存分析的临床试验研究

原文题目：结直肠癌临床病理特征与预后的多因素回归分析

1. 原文主要内容摘要及对存在问题的剖析。

如 22.1 节所示，这篇原文在临床试验设计方面同样存在不少问题，此处不再赘述。下面先复制该文中的一张统计表，见表 22-5；同时，将原文中关于“统计学分析”部分内容复制如下：

应用 SPSS 10.0 软件包进行统计分析。计量资料采用 $\bar{x} \pm s$ 表示，生存率计算采用寿命表法，生存率的比较采用时序检验法。对于单因素分析有意义的变量，通过变量逐步引入的方法进入 Cox 回归模型，从而排除变量之间的干扰因素，确立影响生存预后的独立因素。

表 22-5 761 例结直肠癌患者的病理特征及其对预后的影响(单因素分析)

病理特征	例数	3 年生存率 (%, $\bar{x} \pm s$)	5 年生存率 (%, $\bar{x} \pm s$)	P 值
肿瘤直径 (cm)				
<5	387	63.7 \pm 3.6	60.7 \pm 4.0	>0.05
> =5	374	62.0 \pm 3.4	60.5 \pm 3.6	
肿瘤的大体形态				
肿块型	321	71.0 \pm 3.4	69.4 \pm 3.7	0.0096
溃疡型	318	59.5 \pm 4.1	55.9 \pm 4.6	
侵袭型	122	48.5 \pm 6.4	48.5 \pm 6.4	
肿瘤分化程度				
高分化	195	69.5 \pm 4.3	67.3 \pm 4.7	0.0141
中分化	408	62.3 \pm 3.6	58.7 \pm 4.3	
低分化	158	54.1 \pm 5.6	51.3 \pm 5.5	
肿瘤的组织学				
腺癌	661	67.2 \pm 6.7	64.3 \pm 6.1	>0.05
黏液腺癌 +	100	63.0 \pm 2.8	60.2 \pm 3.1	
肿瘤的病理分期				
A	42	91.1 \pm 6.4	91.1 \pm 6.4	<0.005
B	354	72.3 \pm 3.8	68.2 \pm 4.5	
C	320	58.2 \pm 4.2	56.6 \pm 2.4	
D	45	18.7 \pm 5.8	11.2 \pm 6.8	
肿瘤的肠壁侵袭				
黏膜层	5	100 \pm 0.0	100 \pm 0.0	<0.0005
黏膜下层	18	88.6 \pm 7.6	88.6 \pm 7.6	
肌层	169	69.1 \pm 4.0	65.4 \pm 4.5	
浆膜层	316	67.9 \pm 5.3	64.5 \pm 6.0	
浆膜外	253	49.7 \pm 4.2	49.7 \pm 4.2	
转移情况				
无淋巴转移	396	68.8 \pm 4.0	66.7 \pm 4.4	<0.0005
有淋巴转移	320	64.6 \pm 3.5	64.6 \pm 3.5	
远处转移	45	16.0 \pm 7.5	53.0 \pm 6.7	

由表 22-5 可知：肿瘤的大体形态、肿瘤的病理分期和肿瘤的肠壁侵袭都属于多值名义变量，而肿瘤分化程度属于多值有序变量。对它们如何赋值，原文中未见明确陈述。下面再复制一张反映原文进行多重 Cox 比例风险模型拟合临床试验资料的结果，见表 22-6。

表 22-6 761 例结直肠癌患者临床病理特征多因素 Cox 回归分析结果

变量	偏回归系数	标准误	相对危险度	P 值
大体分型	0.264	0.104	1.302	0.012
分化程度	0.228	0.111	1.256	0.041
肿瘤分期	0.902	0.115	2.464	0.000
浸润深度	0.213	0.109	1.238	0.045

2. 对原文中存在的主要问题的辨析与释疑

原文除了在临床试验设计方面存在很多前文提及的“通病”之外，在拟合多重 Cox 比例风险回归模型的全过程中存在很多致命的错误，现着重论述如下。

(1) 将三个多值名义变量和一个多值有序的自变量简单地赋值 1, 2, 3, ..., 其科学合理性未做证实。

(2) 原文中所分析的生存资料是否满足 PH 假定(比例风险假定), 原文中未做交代, 更没有进行必要的验证。

(3) 基于单因素或单变量的分析结果, 将那些 P 值小于 0.05 的自变量或因素逐一引入 Cox 比例风险模型, 这种筛选自变量的策略是有问题的。因为可能存在那样的自变量: 单个自变量对因变量的影响很小, 但当它们与某些其他自变量同时在回归模型中时, 可能会发挥很大的作用。显然, 这样的自变量就失去了被选中的机会。

(4) 任何人都不能武断地认为: 仅有研究者列出来的那些自变量的一次项与因变量有关系。也可能某些自变量的二次项或某些自变量之间的交互作用项对因变量有重要的影响, 也就是说, 除了原先的那些自变量之外, 还应结合基本常识、专业知识和探索性分析结果, 产生出一些派生自变量, 一同与原先的自变量参与多重回归分析的自变量筛选过程。

(5) 迄今为止, 尚无理由认为某种筛选自变量的方法是绝对最好的。所以, 在创建多重回归方程的过程中, 应尽可能多用几种筛选自变量的方法。

(6) 应创建一套评价最优回归方程优劣的评判标准, 据此从众多较优的回归方程中选择相对最优的一个来下专业结论, 并报告最终结果。以下提供的几条评价标准仅供参考:

- ① 拟合的多重回归方程在整体上有统计学意义。
- ② 多重回归方程中各回归参数(包括截距项)的估计值的假设检验结果都有统计学意义。
- ③ 多重回归方程中各回归参数的估计值的正负号与其后的变量在专业上的含义相吻合。
- ④ 根据多重回归方程计算出因变量的所有预测值在专业上都有实际意义。
- ⑤ 相对来说, 根据多重回归方程计算所得的预测残差平方和最小。
- ⑥ 相比之下, 多重回归方程中自变量的个数较少且便于观测其取值。

(胡良平 吴庆文 陈 娟 贾玉凤 孙日扬 王小莉 王 璐 李 崇)

附录 A 胡良平统计学专著及配套软件简介

1. 《医学统计学与 SAS 应用技巧》简介

胡良平, 周士波主编. 北京: 中国科学技术出版社, 15.67 万字, 1991(定价: 4.2 元).

本书基于 DOS 版 SAS 6.03 软件, 介绍了 SAS 应用入门、医学试验设计、常用统计分析、多元统计分析和 VAX SAS 应用入门。

2. 《医学统计应用错误的诊断与释疑》简介

胡良平主编. 北京: 军事医学科学出版社, 17.8 万字, 1999(定价: 12 元).

本书针对医学科研和医学期刊中常犯的统计学错误, 讲解如何识别错误, 如何正确选用统计分析方法。

3. 《医学统计学内容概要、考题精选与考题详解》简介

胡良平编著. 北京: 军事医学科学出版社, 37 万字, 2000(定价: 22 元).

本书简明扼要地概述了医学统计学的主要内容, 精选出 20 套适合检查统计学应用水平的考题, 并附有详细的解答。

4. 《现代统计学与 SAS 应用》简介

胡良平主编. 北京: 军事医学科学出版社, 1996, 2000, 2002(定价: 40 元).

本书详细地介绍了各种常用和多元统计分析方法, 并给出了手工计算和用 6.04 版 SAS 软件实现统计计算的方法和结果的解释。

5. 《Windows SAS 6.12 & 8.0 实用统计分析教程》简介

胡良平编著. 北京: 军事医学科学出版社, 96.9 万字, 2001(定价: 52 元).

本书不仅介绍了各种常用和多元统计分析方法, 还着重介绍了 Windows SAS 6.12 & 8.0 的使用方法(含编程法和非编程法), 详细介绍了辨析多因素设计类型的技巧和用 SAS 实现实验设计的方法。

6. 《医学统计学基础与典型错误辨析》简介

胡良平, 李子建主编. 北京: 军事医学科学出版社, 60.4 万字, 2003(定价: 36 元).

本书详细地介绍了学习统计学的策略、所必需的基本知识、常用的描述性统计分析方法和假设检验方法。

7. 《检验医学科研设计与统计分析》简介

胡良平主编. 北京: 人民军医出版社, 64 万字, 2004(定价: 65 元).

本书紧紧围绕实验设计的三要素和四原则、分析定量资料和定性资料的要领、诊断性试验和一致性检验中的统计分析方法等重要内容, 从正反两方面详细阐述了学习和灵活运用这些知识的方法和技术。

8. 《医学统计实用手册》简介

胡良平主编. 北京: 人民卫生出版社, 48.5 万字, 2004(定价: 30 元).

鉴于目前医学科研和医学期刊中存在大量误用和滥用统计学的现象,本书通过分析这些现象产生的根源和实质,有针对性地提出了解决这些问题的对策。

9. 《统计学三型理论在实验设计中的应用》简介

胡良平主编.北京:人民军医出版社,50.1万字,2006(定价:45元)。

本书针对“许多人学了多遍统计学仍不得要领,几乎是一用就错”的普遍现象,提出了彻底解决的对策,其精髓就是“统计学三型理论(简称‘三型理论’)”,即统计学问题基本上都可归结为“表现型”、“原型”和“标准型”,准确把握每个具体问题中的“三型”,将能科学合理地解决科研工作中与统计学有关的实际问题。事实上,统计学中的全部内容皆可运用“三型理论”来解说,但本书仅关注“科研设计”,特别是“实验设计”方面的问题。

10. 《医学统计实战练习》简介

胡良平主编.北京:军事医学科学出版社,83.4万字,2007(定价:66元)。

本书收录笔者21年来从事统计教学、科研、咨询和培训工作中积累的各种考试真题以及根据审稿的稿件和公开发表的论文中提取的资料改编而成的新题,总共约有1000余道,并给出了每一道题的详细解答。

11. 《口腔医学科研设计与统计分析》简介

胡良平主编.北京:人民军医出版社,54万字,2007(定价:65元)。

书中给出了取自口腔医学科研设计和统计分析的大量实例,运用“统计学三型理论”辨析“实验设计、统计描述和统计分析”中出现的错误,在给出正确做法的同时,还给出了带有原始数据的各种实例,用SAS软件演示统计分析的全过程和部分手工计算过程。还给出了估计样本含量的公式、实例和用SAS实现计算的方法。

12. 《统计学三型理论在统计表达与描述中的应用》简介

胡良平主编.北京:人民军医出版社,55.3万字,2008(定价:80元)。

本书运用统计学三型理论,透过各种具体的统计表达和描述方面问题的“表现型”,揭示其“原型”,进而将“原型”正确地转变为“标准型”,使统计表达与描述方面的问题尽可能得到圆满解决。

13. 《科研课题的研究设计与统计分析(第一集)》简介

胡良平主编.北京:军事医学科学出版社,72.5万字,2008(定价:55元)。

本书取材于我国2006年500多种生物医学期刊中影响因子较高的23种期刊,查阅这些期刊中近3000篇论著,从中挑选出具有广泛代表性的论著约300篇,主要从统计研究设计和统计分析方法选用两个方面,来剖析论著中存在的统计学问题,从而提示出我国生物医学科研工作质量需要进一步提高。

14. 《医学统计学——运用三型理论分析定量与定性资料》简介

胡良平主编.北京:人民军医出版社,72.3万字,2009(定价:115元)。

本书在统计学思想指导下,运用统计学三型理论,透过各种具体科研问题所呈现的“表现型”,揭示其“原型”,进而将“原型”正确地转变为“标准型”,全面系统地介绍了各种实验设计类型下收集的定量与定性资料的假设检验方法以及用SAS软件实现统计计算和结果解释。除常用的定量与定性资料的统计分析外,还介绍了META分析方法和高维列联表资料的各种处理方法。

15. 《科研课题的研究设计与统计分析(第二集)》简介

胡良平主编.北京:军事医学科学出版社,69.5万字,2009(定价:56元)。

针对科研工作者所写出的学术论文和硕士与博士研究生所写出的学位论文在统计学方面存在很多问题的现实,本书全面介绍了撰写高质量的论文所必须掌握的科研设计知识、统计分析知识和国际著名统计分析系统(SAS 软件)使用知识,并针对生物医学科研领域中一些主干学科的特点,分析了约 15 个主干学科硕士和博士研究生学位论文中存在的统计学错误。从正反两个方面,揭示科研设计和统计分析的重要性,有利于提高科研工作者和研究生的科研素质、科研质量和论文的水平。

16. 《医学统计学——运用三型理论进行多元统计分析》简介

胡良平主编. 北京:人民军医出版社,41.0 万字,2010(定价:70.0 元).

本书涵盖了现代多元统计分析方法中的绝大部分内容,运用三型理论为指导,对多元统计分析方法进行了科学的分类,有利于实际工作者学习和使用。其内容包括变量聚类分析、主成分分析和探索性因子分析、典型相关分析、结构方程模型分析、无序样品聚类分析和有序样品聚类分析、多维尺度分析、各种设计定量资料的多元方差分析和多元协方差分析、判别分析、对应分析及其 SAS 实现。

17. 《心血管病科研设计与统计分析》简介

胡良平主编. 北京:人民军医出版社,47.5 万字,2010(定价:60 元).

本书内容分正反两个方面,正面讲述统计学中的主要内容,包括统计表达与描述、实验设计、定量与定性资料统计分析、简单相关回归分析和多重回归分析;围绕这些内容,又针对人们误用统计学的实际案例,对差错进行辨析与释疑。无论是正面还是反面内容,基本上都取材于与心血管疾病有关的我国数十种学术期刊中的科研论文。

18. 《SAS 统计分析教程》简介

胡良平主编. 北京:电子工业出版社,106.5 万字,2010(定价:68 元).

本书内容丰富且新颖,实用面宽且可操作性强。涉及定量与定性资料差异性和预测性分析、变量间和样品间相互与依赖关系及近似程度分析、数据挖掘与基因表达谱分析、绘制统计图与实验设计、SAS 语言和 SAS 非编程模块用法。这些内容高质量、高效率地解决了实验设计、统计表达与描述、各种常用和多元统计分析、现代回归分析和数据挖掘、SAS 语言基础和 SAS 实现及结果解释等人们迫切需要解决却又十分棘手的问题。

19. 《SAS 实验设计与统计分析》简介

胡良平主编. 北京:人民卫生出版社,88.8 万字,2010(定价:72 元).

本书内容涉及面十分宽泛,由 SAS 软件基础、SAS 非编程模块介绍、SAS 编程法用法介绍、SAS 高级编程技术及其应用和 SAS 语言基础这 5 篇组成,涵盖了 SAS 软件及其语言的基础和高级用法,实验设计、统计表达与描述和统计分析的主要内容及 SAS 实现。

20. 《医学统计学——运用三型理论进行现代回归分析》简介

胡良平主编. 北京:人民军医出版社,45.2 万字,2010(定价:75 元).

本书介绍了现代回归分析方法中的大部分内容,包括多重线性回归分析、岭回归分析、各种复杂曲线回归分析、主成分回归分析、Poisson 回归分析、Probit 回归分析、负二项回归分析、配对和非配对设计定性资料多重 logistic 回归分析、对数线性模型分析、生存分析和时间序列分析。

21. 《医学遗传统计分析与 SAS 应用》简介

胡良平,郭晋主编. 北京:人民卫生出版社,41.3 万字,2011(定价:36 元).

本书结合实例,介绍了如何用 SAS 实现四大类遗传数据的统计分析方法,并介绍了简明遗传学的概念与原理、遗传资料统计分析的原理。

22. 《正确实施科研设计与统计分析——统计学三型理论的应用与发展》简介

胡良平主编.北京:人民军医出版社,87.8 万字,2011(定价:139 元).

本书全面介绍了如何在三型理论指导下进行科研设计、统计表达与描述、常用统计分析、现代回归分析、多元统计分析和 SAS 实现方法。科研设计部分涵盖了概念、要点、设计类型等;统计表达与描述部分涵盖了统计表、统计图和概率分布等;常用统计分析部分涵盖了一元定量与定性资料的差异性分析;现代回归分析部分涵盖了包括多重线性回归分析、生存分析和时间序列分析等十余种现代回归分析方法;多元统计分析部分涵盖了包括变量聚类分析、判别分析和对应分析等十余种现代多元统计分析方法;以上各部分均涉及如何用 SAS 软件巧妙实现的技术和方法,并有配套软件 SASPAL 方便程序调用。

23. 《中医药科研设计与统计分析》简介

胡良平,王琪主编.北京:人民卫生出版社,41.4 万字,2011(定价:36.00 元).

本书结合中医药领域中科研实例,不仅从正面介绍了试验设计、统计表达与描述、统计分析方法及 SAS 实现技术,还对实际工作者在运用前述内容过程中所犯的各种错误进行了辨析与释疑。

24. 《临床科研设计与统计分析》简介

胡良平,陶丽新主编.北京:中国中医药出版社,70.7 万字,2012(定价:45.00 元).

本书主要对临床科研设计与统计分析问题进行阐述,同时还用较大篇幅揭示了临床科研课题和论文中的统计学错误,并给出了辨析与释疑。全书中的统计计算均用 SAS 软件实现。

25. 《面向问题的统计学——(1) 科研设计与统计分析》简介

胡良平主编.北京:人民卫生出版社,119.1 万字,2012(定价:98.00 元).

本书分为 6 篇共 54 章,内容涉及消除学习统计学时的心理顾虑、统计思想、三型理论、科研设计、质量控制、表达与描述、单因素设计一元定量与定性资料统计分析、单组设计二元定量资料相关与回归分析和 SAS 语言基础与高级编程技术。

26. 《面向问题的统计学——(2) 多因素设计与线性模型分析》简介

胡良平主编.北京:人民卫生出版社,97.5 万字,2012(定价:80.00 元).

本书分为 6 篇共 52 章,内容涉及多因素试验设计类型及其定量与定性资料的差异性分析和现代回归分析、判别分析、生存分析和时间序列分析;还介绍了多水平模型分析法和综合分析法。

27. 《面向问题的统计学——(3) 试验设计与多元统计分析》简介

胡良平主编.北京:人民卫生出版社,85.2 万字,2012(定价:65.00 元).

本书分为 5 篇共 25 章,内容涉及三类典型的多元数据结构(单组设计多元定量资料、单因素多水平设计多元定量资料、相似或不相似度矩阵)的各种多元统计分析方法,其代表性方法有主成分分析、样品聚类分析、对应分析、多维尺度分析、多元方差和协方差分析。

28. 《外科科研设计与统计分析》简介

胡良平,毛玮主编.北京:中国协和医科大学出版社,40 万字,2012(定价:38 元).

本书分为三篇,第 1 篇为统计学内容概要,包括统计表达与描述、试验设计、定量与定性资料的统计分析、简单相关与回归分析、多重线性回归分析与多重 logistic 回归分析;第 2 篇为外科科研

中常见统计学错误辨析与释疑；第3篇为医学统计学要览，以“问题引导”的形式提纲挈领地介绍了“科研设计要览”与“统计分析要览”。

29.《科研设计与统计分析》简介

胡良平主编.北京：军事医学科学出版社，130.5万字，2012(定价：98元)。

本书用7篇31章130多万字的篇幅，概述了国内外迄今为止应该涵盖在统计学之中的绝大部分精彩内容：富含唯物辩证法精髓和心理学分析的统计思想、使统计思想具体化并具有可操作性的三型理论、灵活运用三型理论解决科研设计、统计表达与描述、各种简单与复杂统计分析、用国际著名统计分析系统SAS实现与前述全部内容有关的计算、结果解释和结论陈述。

30.《呼吸系统科研设计与统计分析》简介

胡良平，鲍晓蕾主编.北京：军事医学科学出版社，53.8万字，2013.1(定价：55元)。

本书以近几年出版的与呼吸科研相关杂志为主要的资料来源，在阐述统计学的基本理论、知识和技能的基础上，突出培养统计学思维方法、科研设计能力和应用统计分析方法的能力，以及计算机在处理临床科研资料中的正确应用技术。书中还用较大篇幅介绍了呼吸科研课题和论文中常见统计学错误案例的辨析与释疑、SAS软件的基础知识和使用技巧。

31.《护理科研设计与统计分析》简介

胡良平，关雪主编.北京：军事医学科学出版社，47.7万字，2013.1(定价：50元)。

本书以近几年出版的护理科研相关杂志为主要的资料来源，在阐述统计学的基本理论、知识和技能的基础上，突出培养统计学思维方法、科研设计能力和应用统计分析方法的能力，以及计算机在处理护理科研资料中的正确应用技术。书中自始至终采用“识别错误”、“正确引导”和“归纳总结”的写作思路，把人们实施护理课题和撰写学术论文中常出现的错误呈现出来，并逐一对错差进行辨析与释疑；对案例所涉及的统计学基础知识进行系统梳理，从正面加以引导；对有关的统计理论和方法，从原理上进行归纳总结，以便实际工作者不仅知其然，还能知其所以然。

32.《脑血管病科研设计与统计分析》简介

胡良平，贾元杰主编.北京：军事医学科学出版社，50.3万字，2013.5(定价：58元)。

本书结合脑血管病临床科研实际，比较全面地介绍了从事临床科研工作所必需的思维方法、统计学基础理论和基本的统计分析技术，内容包括统计思想与三型理论在脑血管病科研中的应用、脑血管病科研基础——统计表达与描述、脑血管病科研设计、脑血管病试验设计、脑血管病临床试验设计、脑血管病调查设计、样本量估计与检验效能分析、常见多因素试验设计类型辨析、定量与定性资料统计分析、简单相关与回归分析、多重线性回归分析与多重logistic回归分析。

33.《临床试验设计与统计分析》简介

胡良平，陶丽新主编.北京：军事医学科学出版社，54.7万字，2013.5(定价：58元)。

本书结合临床科研和临床试验实际，首先介绍了临床前研究和临床研究的主要内容，不仅在内容的安排上达到了承上启下的效果，使读者很自然地进入临床试验的情境之中；而且在要点把握上也起到了言简意赅、纲举目张的作用，使读者能在尽可能短的时间内领悟和抓住临床试验的核心和要领。在此基础上，结合笔者在国家级新药评审中发现的诸多问题，揭示了新药或医疗器械临床试验研究中的陷阱和识别错误的策略；介绍了如何把握好临床试验研究中的三要素、四原则、设计类型和比较类型的概念、方法和技术要领。进而针对临床试验研究中使用频率最高的设计类型——成组设计，围绕四种比较类型、定量与定性资料、假设检验、样本量和检验效能估计等关键性问题，结合临床实例逐一进行介绍，并对同类问题进行了比较研究。

34.《非线性回归分析与 SAS 智能化实现》简介

胡良平, 高辉主编. 北京: 电子工业出版社, 51.5 万字, 2013.7(定价: 39 元).

本书概述了回归分析的概念、分类、简单直线、曲线回归分析和多重线性回归分析、复杂固定模式和非固定模式曲线回归分析、单水平和多水平多重曲线回归分析。每种回归分析方法, 都介绍了分析目的、数据结构(问题与数据)、切入点(分析与解答)、统计模型(计算原理)、分析步骤(含 SAS 实现)。在固定模式单水平非线性回归分析中, 涉及的统计模型有二项型和三项型指数曲线模型、logistic 和 Gompertz 和 Richards 生长曲线模型、Bleasdale-Nelder 和 Halliday 和 Farazdaghi-Harris 产量-密度曲线模型; 在非固定模式单和多水平多重非线性回归分析中, 涉及的统计模型有二值结果变量定性资料单和多水平 logistic 和 probit 和互补双对数回归模型; 多值有序结果变量定性资料单和多水平累积 logistic 和 probit 和互补双对数回归模型; 多值名义结果变量定性资料单和多水平扩展 logistic 回归模型和计数资料单和多水平 Poisson 和负二项回归模型。在上述各种情况下, 还给出了同类问题的比较研究和 SAS 智能化实现及结果解释。

35.《课题设计与数据分析——关键技术与标准模板》简介

胡良平主编. 北京: 军事医学科学出版社, 48.5 万字, 2014.1(定价: 48 元).

本书以“如何做好科研课题”为出发点和落脚点, 开门见山, 直奔主题。第 1 章从正反两种不同的视角, 全面介绍了课题设计的基本概念、关键技术、具体做法和常见错误的辨析与释疑; 第 2 章介绍了智源临床研究执行平台, 它是一个智能化很高的数据管理和数据分析软件平台, 它集数据网络平台录入、随机分组、逻辑核查、与国际著名统计分析软件 SAS 实现无缝对接等功能于一身; 第 3 章和第 4 章介绍了临床试验研究中不可缺少的两个关键技术, 即样本含量估计和随机化的 SAS 实现; 第 5 章介绍了临床试验数据管理的标准操作规程、质量控制、具体流程和建立数据库的多款软件; 第 6 章到第 9 章介绍了与高质量完成科研课题密切相关的第二部分的内容, 即对资料的统计表达描述和各种统计分析。与众不同的是, 书中所介绍的统计分析方法几乎都可采用 SAS 智能化实现, 免去了使用者在分析过程中很多不必要的担心和麻烦。

36.《SAS 语言基础与高级编程技术》介绍

胡良平, 胡纯严主编. 北京: 电子工业出版社, 73.3 万字, 2014.5(定价: 59 元).

本书全面系统地介绍了国际著名的统计分析系统 SAS 软件的主要内容, 包括 SAS 语言基础、SAS 高级编程技术、SAS 9.2 和 9.3 版新增内容及用法简介、用 SAS 实现试验设计及处理病态数据的两个过程简介, 其中前两部分是本书的重点。SAS 语言基础部分涵盖了如下内容: SAS 软件介绍、导入访问外部数据、基本 SAS 语言及其应用、常用 SAS 函数及其应用; 而 SAS 高级编程技术部分包括如下内容: 宏及其应用、SQL 及其应用、ODS 及其应用、数组 (ARRAY) 及其应用、IML 及其应用和如何掌握 SAS 语言的核心技术。书中还介绍了 SAS 9.2 和 9.3 版中一些新增过程和新增选项以及部分实用新过程的使用方法和技巧。

37.《医学综合统计设计与数据分析》介绍

胡良平, 赵铁牛, 李长平主编. 北京: 电子工业出版社, 59.8 万字, 2014.5(定价: 45 元).

本书作者查阅了医学综合类期刊数十种, 以每种期刊近 5 年中刊载的学术论文为研究对象, 概括地总结了各类期刊论文中统计学应用的现况和在统计学应用方面存在的问题, 先从正面讲解在生物医学和临床科研领域内如何正确使用统计学的理论、技术和方法, 又列举了各种期刊论文中经常出现的误用和滥用的典型案例, 对案例中出现的差错进行辨析与释疑, 从正反两个方面启发和引导广大读者正确运用统计学、提高识别统计学错误的能力。内容涉及统计设计、统计表达与描述、

统计分析、SAS 实现、结果解释和结论陈述。最后，又以一章的篇幅，提纲挈领、言简意赅地总结了学习统计学的方法。

38.《基础医学统计设计与数据分析》介绍

胡良平, 余红梅, 高辉主编. 北京: 电子工业出版社, 50.56 万字, 2014.6(定价: 45 元).

本书作者查阅了基础医学和医学综合类期刊数十种, 以每种期刊近 5 年中刊载的学术论文为研究对象, 概括地总结了各类期刊论文中统计学应用的现况和在统计学应用方面存在的问题, 先以 10 章的篇幅列举了各种期刊论文中时常出现的误用和滥用的典型案例, 对案例中出现的差错逐一进行辨析与释疑, 从反面引起广大读者的警觉, 目的是用“反面刺激”有效提高广大读者识别统计学应用中可能出现的各种错误的能力, 以期达到“吃一堑长一智”之学习功效; 接着又以 5 章的篇幅从正面比较系统地介绍了统计学的核心内容, 内容涉及科研设计、统计表达与描述、统计分析方法合理选择和 SAS 实现等关键技术, 以“提纲挈领、言简意赅”的写作手法, 将统计学本身和使用方法的精华呈现在读者的面前。

39.《实用医学统计学》介绍

胡良平主编. 北京: 金盾出版社, 22.6 万字, 2014.8(定价: 22.0 元).

本书旨在全面介绍医学统计学的基础知识, 内容涉及两篇, 即第 1 篇医学统计学基础和第 2 篇单组设计及其资料的统计处理。第 1 篇包括以下 7 章, 即试验设计基础(重点讲述其核心技术, 即试验设计三要素、四原则、设计类型和质量控制)、统计分析基础与工具(重点讲述统计分析方法概述、SAS 软件简介、SAS 用法简介和 SASPAL 简介)、试验设计关键技术(估计样本含量和进行随机化的方法)、统计资料与基本概念、相对数与分位数、平均指标与变异指标、统计表与统计图; 第 2 篇包括以下 4 章, 即与单组设计有关的概念问题、与单组设计有关的区间估计问题、与单组设计有关的假设检验问题和与单组设计有关的相关与回归分析问题。上述内容不仅是医学统计学的重要基础, 而且, 在写作手法上, 注重从实际出发, 便于学习者易学易懂易用。

40.《新概念统计学——统计思维与三型理论在医药领域中的应用》介绍

胡良平, 张天明主编. 北京: 中国中医药出版社, 17.9 万字, 2015.2(定价: 29.0 元).

本书以统计思维和三型理论为一条红线, 贯穿始终。从此理论开始发挥作用的统计学学科入手, 延伸了触角、拓展了视野, 将其应用于以医学为主要学科的众多研究领域, 涉猎的范围包括统计学、医学、生物学、心理与认知科学。通读本书, 可以使具有不同文化素养和知识结构的人都会受益匪浅, 因为本书是来自军事医学科学院和中国科学院大学多位硕士研究生关于医药科学的思考而写出的心得, 也是彰显他们才华和智慧的一个缩影。用统计思维和三型理论去解读与医药科研有关的问题, 不仅可以在认识和解事物和现象时达到“化繁为简、由表及里”之目的, 甚至可以产生“洞见症结、醍醐灌顶”之功效。

41. 配套软件简介

(1)与《现代统计学与 SAS 应用》和《Windows SAS 6.12 & 8.0 实用统计分析教程》两本书对应的 SAS 引导程序, 即 SASPAL 软件由李子建研制, 需要者可发电子邮件联系: lphu812@sina.com。

(2)与《口腔医学科研设计与统计分析》一书对应的 SAS 引导程序, 即 SASPAL 软件由胡纯严研制, 需要者可发电子邮件联系: valencia@sina.com 或 lphu812@sina.com。

(3)与《统计学三型理论在统计表达与描述中的应用》一书对应的 SAS 引导程序, 即 SASPAL 软件由胡纯严研制, 需要者或有疑问者请发电子邮件联系: valencia@sina.com。

(4)与《统计学三型理论在定量与定性资料统计分析中的应用》一书对应的 SAS 引导程序, 即

SASPAL 软件由胡纯严研制，需要者或有疑问者请发电子邮件联系：valencia@sina.com。

(5) 与《医学统计学——运用三型理论进行多元统计分析》一书对应的 SAS 引导程序，即 SASPAL 软件由胡纯严研制，需要者或有疑问者请发电子邮件联系：valencia@sina.com。

(6) 与《医学统计学——运用三型理论进行现代回归分析》一书对应的 SAS 引导程序，即 SASPAL 软件由胡纯严研制，需要者或有疑问者请发电子邮件联系：valencia@sina.com。

(7) 与《正确实施科研设计与统计分析——统计学三型理论的应用与发展》一书对应的 SAS 引导程序，即 SASPAL 软件由胡纯严研制，需要者或有疑问者请发电子邮件联系：valencia@sina.com。

(8) 与《SAS 统计分析教程》一书对应的 SAS 引导程序，即 SASPAL 软件由胡纯严研制，需要者或有疑问者请发电子邮件联系：valencia@sina.com。

(9) 与《科研设计与统计分析》一书对应的 SAS 引导程序，即 SASPAL 软件由胡纯严研制，需要者或有疑问者请发电子邮件联系：valencia@sina.com。

(10) 与《非线性回归分析与 SAS 智能化实现》一书对应的 SAS 引导程序，即 SASPAL 软件由胡纯严研制，需要者或有疑问者请发电子邮件联系：valencia@sina.com。